Big Data Analytics towards a Retrofitting Plan for the City of Stockholm

Bram van der Heijde

Master of Science Thesis
Stockholm 2014

Bram van der Heijde

Big Data Analytics towards a Retrofitting Plan for the City of Stockholm

Supervisor:
Hossein Shahrokni
Examiner:
Nils Brandt

Master of Science Thesis
STOCKHOLM 2014

PRESENTED AT
INDUSTRIAL ECOLOGY
ROYAL INSTITUTE OF TECHNOLOGY

TRITA-IM-EX 2014:08

Industrial Ecology,
Royal Institute of Technology
www.ima.kth.se

"Es ist nicht genug, zu wissen, man muß auch anwenden;
es ist nicht genug, zu wollen, man muß auch tun."

Johann Wolfgang von Goethe

Abstract

This thesis summarises the outcomes of a Big Data analysis, performed on a set of hourly district heating energy consumption data from 2012 for nearly 15 000 buildings in the City of Stockholm. The aim of the study was to find patterns and inefficiencies in the consumption data using KNIME, a big data analysis tool, and to initiate a retrofitting plan for the city to counteract these inefficiencies. By defining a number of energy saving scenarios, the potential for increased efficiency is estimated and the resulting methodology can be used by other (smart) cities and policy makers to estimate savings potential elsewhere. In addition, the influence of weather circumstances, building location and building types is studied.

In the introduction, a concise overview of the concepts Smart City and Big Data is given, together with their relevance for the energy challenges of the 21st century. Thereafter, a summary of the previous studies at the foundation of this research and a brief theory review of less common methods used in this thesis are presented.
The method of this thesis consisted of first understanding and describing the dataset using descriptive statistics, studying the annual fluctuations in energy consumption and clustering all consumer groups per building class according to total consumption, consumption intensity and time of consumption. After these descriptive steps, a more analytical part starts with the definition of a number of energy saving scenarios. They are used to estimate the maximal potential for energy savings, regardless of actual measures, financial or temporal aspects.

This hypothetical simulation is supplemented with a more realistic retrofitting plan that explores the feasibility of Stockholm's Climate Action Plan for 2012-2015, using a limited set of energy efficiency measures and a fixed investment horizon. The analytical part is concluded with a spatial regression that sets out to determine the influence of wind velocity and temperature in different parts of Stockholm.

The conclusions of this thesis are that the potential for energy savings in the studied data set can go up to 59% or 4.6 TWh. The financially justified savings are estimated at ca. 6% using favourable investment parameters. However, these savings quickly diminish because of a high sensitivity on the input parameters. The clustering analysis has not yielded the anticipated results, but they can be used as a tool to target investments towards groups of buildings that have a high return on investment.

Keywords: Smart City, Big Data, Energy efficiency, District heating, Stockholm

Acknowledgement

Over the past few months, I have learnt that a thesis research is not made on one's own. Therefore, I would like thank everyone who has made this thesis possible, be it directly or indirectly.

In the first place, I owe a lot of thanks to Hossein Shahrokni. Without him, I would not have written a thesis about this particular subject. He has always inspired me with what to research next and motivated me to dig deeper and push my boundaries. I have to thank Hossein for introducing me to the world of conferences and publishing scientific papers.
Next, I want to thank Fabian Levihn, who has been an indispensable team member in writing the paper that is connected to this thesis. His advice on how to analyse the economic aspects of the saving scenarios was very important.
Thanks to Jan Haas, I could dispose of the shapefiles of Stockholm's zip code areas. It is no exaggeration to state that without him, there would be no maps in this thesis.

Further, many thanks go to EIT KIC InnoEnergy, who made it possible for me to spend a year in Sweden and get a dual degree. I would especially like to acknowledge professor Johan Driesen and Mar Martinez Diaz for everything they have done for us.

Next, I would like to show my appreciation to my friends Wim and Yixiao. Our discussions about our research have improved my thesis in many ways. Apart from the serious stuff, I had a great time with you in Sweden, not to mention our many visits to our favourite Chinese restaurant. I want to thank Christophe as well, for joining our "smart cities thesis team" and for making valuable additions to our joint results.

Last but not least, I want to thank my family. My parents, because they have always helped me and supported me. I realise that without them, I would not be where and what I am now. And Lieselotte, thank you for being you. Thanks for having patience with my absence for an entire year.

Stockholm, May 2014

Contents

List of Figures

List of Tables

Abbreviations

CFC	**C**hloro**f**luoro**c**arbon
DH	**D**istrict **H**eating
EUI	**E**nergy **U**se **I**ntensity
GHG	**G**reen**h**ouse **G**as
GIS	**G**eographic **I**nformation **S**ystem
HRV	**H**eat **R**ecovery **V**entilation system
ICT	**I**nformation and **C**ommunications **T**echnology
IQR	**I**nter**q**uartile **R**ange
ROI	**R**eturn **O**n **I**nvestment
VHR	**V**entilation **H**eat **R**ecovery

Symbols

E	energy	MWh ($3600 \cdot 10^6$ J)
EUI	energy use intensity	MWh m^{-2}, unless stated otherwise
EUI^*	dimensionless EUI	
h	heat transfer coefficient	W K^{-1}m^{-2}
k	thermal conductivity	W m^{-1}K^{-1}
L	distance	m
P	power	W (Js^{-1})
Q	heat flow	W
T	temperature	$^{\mathrm{o}}$C

To my mother and my father

Chapter 1

Introduction

1.1 Background

In order to understand this master thesis, it is important to have an insight in the background against which the research took place. This section describes the larger picture of challenges that make the study of energy efficiency necessary, and on the other hand technological developments that enabled this study to be conducted.
Smart Cities and *Big Data* are well-known buzz words in this field of study. However, you – the reader – might not be too familiar with these concepts; therefore, a short introduction is provided as well.

1.1.1 Challenges for the 21st century

The history of Man has been characterized by the discovery and improvement of countless techniques that improved the quality of life in some way. But together with these discoveries, population grew and the consumption of the Earth's resources increased steadily. In the course of the 20th century, the awareness that there are limits to this growth (Meadows et al., 1972) started to develop. The difficulties and challenges attached to these growth limits can be summarized in the keywords below.

Population growth and urbanisation The world population is growing exponentially. In addition to this population growth, more and more people leave rural areas and move towards cities, a phenomenon otherwise known as urbanisation (WHO, 2013). According to the United Nations (2012), the world population increased from 6.1 billion in 2000 to 6.9 billion in 2010. Nowadays, the number has already surpassed 7.1 billion (US Census Bureau, 2014) and following current projections, world population is estimated over 8 billion in 2025.

In addition, the United Nations (2012) also provides detailed information about the urban and rural population. Indeed, the proportion of people living in cities had just surpassed 50% in 2010, and is predicted to reach 60% by 2030.
On the one hand, the continuously growing world population makes the consumption of raw material and energy sources from the world grow at more or less the same rate (see the Kaya and IPAT identities in Waggoner and Ausubel (2002)). At the same time, the increasing proportion of people living close together in cities increases the difficulty of providing services, energy and goods to the population, and managing the corresponding waste streams.

Sustainability The next question that is introduced by the population growth, is whether the growing consumption and waste production can be *sustained* or not. But what is sustainability? Ehrenfeld (2004) defines the concept as *"the possibility that human and other forms of life will flourish on the planet forever"*. However, this definition is only partly conclusive, mostly because of the uncertainty that comes with time.
As Graedel and Allenby (2010) explain, the understanding of sustainability depends largely on the time scale that is chosen. Indeed, on a shorter time scale anything can be sustainable, since there is no need to worry about feedback effects from waste streams or the depletion of resources; on the other hand, on the very long time scale, no system is sustainable because of the continuous increase of entropy (the Second Law of Thermodynamics (Moran et al., 2011), closely related to the Law of Conservation of Misery). In addition to the time boundary, also the spatial boundary of the system influences its sustainability.
Independent of the discussion about the definition of sustainability, it is generally agreed upon that the current course of increasing consumption is not sustainable. Apart from the increasing population, the most imminent danger to our planet's eco-system is the global warming phenomenon or climate change. This problem is the subject of the next paragraph.

Energy and climate change As described by IPCC (2014), the phenomenon of global warming is directly related to the concentration of greenhouse gases (GHG) in the atmosphere. The lion's share of these gases (mostly CO_2, but also methane and CFCs) are caused by human activities. In the case of carbon dioxide, the emissions are mainly caused by the combustion of fossil fuels for energy purposes.
In order to decrease the amount of GHG emissions, a transition in the *generation* and

Remark that according to the First Law of Thermodynamics, energy cannot be generated nor destroyed, but only converted from one form to another and hence, the use of the words *generation* and *consumption* is not correct. However, for the sake of simpler formulation, these words will be used to indicate that energy is being transformed from a raw energy material to a form that is useful for the consumer.

consumption of energy is needed. In the case of generation, a replacement of the current fossil fuel based system towards renewable and low-emission energy sources is necessary. The concept of smart cities can support this transition to more intermittent (and thus less reliable) energy sources by the implementation of demand side management (DMS). In the second case, that of consumption, a decrease in emissions can be established by increasing the energy efficiency of various types of consumption.

These two effects, together with the influence of population increase, are illustrated by the Kaya identity (IPCC, 2014; Waggoner and Ausubel, 2002). This relation is almost trivial, but succeeds very well in showing the influence of all mentioned factors.

$$I_{CO_2} = P \, \frac{GDP}{P} \, \frac{E_{cons}}{GDP} \, \frac{I_{CO_2}}{E_{cons}}, \tag{1.1}$$

with I_{CO_2} the global CO_2 emission, P the world population, GDP the global gross domestic product and E_{cons} the energy consumed. If the terms in fractions are gathered in one factor, the influence factors become clear:

$$I_{CO_2} = P \; g \; e \; i, \tag{1.2}$$

where g is the global GDP per capita, e the factor that denotes energy needed to create a certain added value, and i the emissions of CO_2 per amount of energy that is generated. This equation shows the amount of emissions for a given population, economic situation, technology and energy intensity. But what's even more interesting, is to investigate the rate of change for this equation.

In this thesis, the main emphasis is on the e term, which can be interpreted as the total energy needed to provide a service, in this case the heating of buildings. The current research tries to provide insights in how the district heating consumption is established on the one hand (the value of e, who is consuming what), and investigates the potential for savings on the other hand (the rate of change of e).
Of course, energy is not only consumed by buildings; the composition of energy end-uses has been studied by the IPCC (2014), and is shown in figure 1.1. Apparently, buildings constitute the largest portion of the final energy consumption with a share of 34%. This leads to conclude that the decrease of building energy use intensity can have a large influence on the total energy use, GHG emissions and finally climate change.

1.1.2 Smart cities

Although (or rather because) there has been a lot of research on the topic of smart cities, a plethora of varying definitions for what a smart city is can be found. However, in most of the cases the *smartness* of the cities comes from the utilization of ICT.

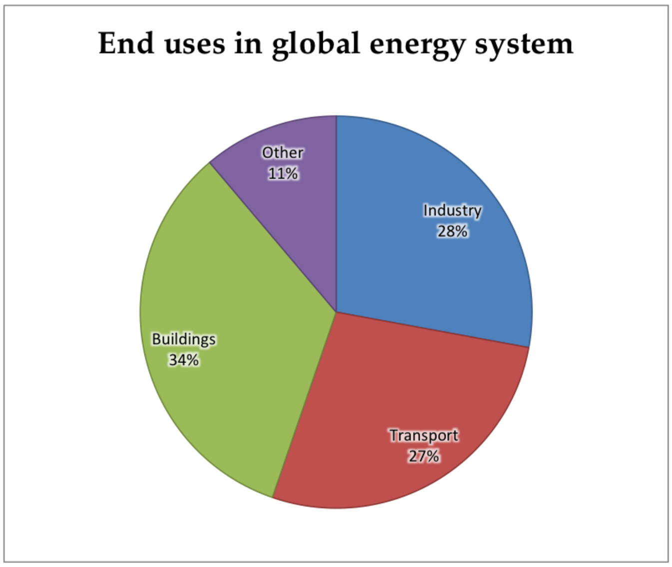


FIGURE 1.1: End uses in global energy system as of 2010, adapted from IEA (2012)

Hollands (2008) argues that in addition to smartness and connectivity through ICT, also smartness in sociological facets of the city is needed. In this train of thought, education, culture, politics, economy and even creativity are mentioned; the smart city concept is thus not only about technological capital in a city, but also the so-called *human capital*, or the stock of knowledge and competences present in a city.
Finally, an always recurring aspect of smart cities is the goal of making cities sustainable while at the same time maintaining a healthy economic growth and a high quality of life. Indeed, as mentioned in previous section, the constantly growing urban population poses extensive difficulties on the way life is organised in cities.
Caragliu et al. (2011) summarize all aforementioned aspects in one comprehensive definition:

> "[...] in a smart city, investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance."

In spite of this circumscription of the smart city concept, there is a need for a way to decide whether a city is smart or not. Therefore, the EU has tried to construct a set of criteria which a city has to fulfill in order to be called a Smart City (Lazaroiu and Roscia, 2012). The six "smart" aspects that are assessed are economy, mobility, people, environment, living and governance. All of these aspects are subdivided in more specific criteria, but it is not within the scope of this brief introduction to smart cities to study these in more depth.

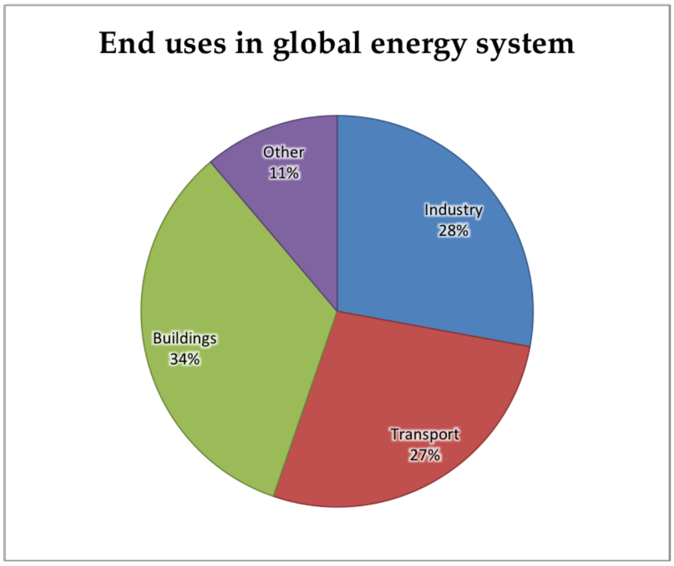


FIGURE 1.1: End uses in global energy system as of 2010, adapted from IEA (2012)

Hollands (2008) argues that in addition to smartness and connectivity through ICT, also smartness in sociological facets of the city is needed. In this train of thought, education, culture, politics, economy and even creativity are mentioned; the smart city concept is thus not only about technological capital in a city, but also the so-called *human capital*, or the stock of knowledge and competences present in a city.
Finally, an always recurring aspect of smart cities is the goal of making cities sustainable while at the same time maintaining a healthy economic growth and a high quality of life. Indeed, as mentioned in previous section, the constantly growing urban population poses extensive difficulties on the way life is organised in cities.
Caragliu et al. (2011) summarize all aforementioned aspects in one comprehensive definition:

> "[...] in a smart city, investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance."

In spite of this circumscription of the smart city concept, there is a need for a way to decide whether a city is smart or not. Therefore, the EU has tried to construct a set of criteria which a city has to fulfill in order to be called a Smart City (Lazaroiu and Roscia, 2012). The six "smart" aspects that are assessed are economy, mobility, people, environment, living and governance. All of these aspects are subdivided in more specific criteria, but it is not within the scope of this brief introduction to smart cities to study these in more depth.

1.1.3 Big data

Returning to the use of ICT in smart cities, the digital revolution has enabled the channelling and storage of increasingly large amounts of data. This handling of larger and larger amounts of data can be characterised by three V's: volume, velocity and variety (Cukier and Mayer-Schoenberger, 2013). New technologies allow data to be transmitted and stored at very high speeds (optical cable, solid-state drives), high quantities (petabytes and onwards) and additionally in a great variety of formats.

However, the concept of big data does not end with the three V's: the term also encompasses the analysis of the data. Hey et al. (2009) explain that this use of big data can be seen as a *Fourth Paradigm* in research. They explain that by exploring patterns in enormous amounts of data using data management and statistics, scientific discoveries can be made. This in contrast to the first three paradigms, viz. empirical observation (experimental science), theoretical science (models and laws) and computational science (in which increasingly complex models are solved numerically instead of analytically).

When the concepts of big data and the fourth research paradigm are brought together with the concept of smart cities, much more insight can be gained in the processes that take place within a city, such as energy provision, services waste streams and even human behaviour, and that are otherwise invisible. This knowledge allows various stakeholders to make better decisions in order to improve these processes and their efficiency. In the end, this optimisation leads to the achievement of the large goal of smart cities, namely sustainability.

1.2 Aim and objectives

1.2.1 Aim

The aim of this master thesis is to analyze big data sets for district heating energy consumption in the City of Stockholm, in order to support the city in the construction of a retrofitting plan. The analysis sets out to gain insight in the consumption data and current inefficiencies, and to consumers by their consumption behaviour. These results will be used to assess the energy saving potential and to construct a cost-efficient retrofitting plan for the city.

1.2.2 Objectives

The objectives are the following:

- To understand the input data,

- to curate and explore the data and combine them into a database,
- learning to work with the KNIME analysis tool and preparing the data for this analysis,
- grouping the energy use data and analyzing the resulting clusters;
- finding underlying patterns (e.g., weather, location,. . .) in the consumption data, and
- identifying inefficiencies and proposing a preliminary retrofitting plan.

1.3 Outline

The next section will give a concise overview of previous studies that are related to the subject of this thesis. In the theory subsection, necessary background information about the theory of clustering is given.

The methodology section firstly provides information about the used data sets and the analysis tools that were utilised. It continues to explain how the data was initially processed and how outliers were removed. A short introduction to the production of energy maps is given, after which the grouping of buildings in six classes and the calculation of energy use intensity is explained. Thereafter, the clustering parameters are determined. The savings scenarios and measures for the maximal savings potential analysis and the retrofitting analysis are specified and finally, the regression analysis to investigate the weather influence is explained.

The results section presents the results of the methods from the previous section in the same order. However, in the first place the descriptive statistics of the data set are summarised. The discussion section provides additional contemplations about the interpretation of the results.

Chapter 2

Theory and previous work

2.1 Previous work

This section summarises some of the sources that were consulted to develop the currently used analysis methods. Although not many predecessors in this field of study were found, a number of interesting similar research projects were encountered. Other studies provide more information about the situation of heating energy consumption in Sweden.

2.1.1 Electricity consumption analysis for Ireland

The work of Rosaria Silipo and Phil Winters was the main inspiration for this thesis research. Their white paper "Big Data, Smart Energy and Predictive Analysis – Time Series Prediction of Smart Energy Data" (Silipo and Winters, 2013) studied smart energy data from the Smart Energy Trials in Ireland. The data set comprises half-hourly electricity values for 6000 houses and businesses. In this paper, KNIME was used for all data manipulation and calculation steps. All steps are described meticulously with the required actions in KNIME and is thus an excellent guide to using this program for big data analysis.

The aim of this paper was twofold. The first was to identify clusters containing different consumer groups. The second was to predict consumption data using these clusters and algorithms based on autocorrelation.

The first step that is described in the white paper is the importation and transformation of the electricity data. The consumption figures are aggregated on different time scales. At the same time, the proportion of consumption on daily and weekly basis is calculated for each smart meter.

The results of this initial transformation step were inserted in a k-means clustering algorithm. The clustering was based on the following variables: percentage values for the proportion of consumption on each week and weekend day and each hour of the day, average consumption per hour, day, week, month and year, as well as the average consumption on week days and weekend days and the total energy consumption over the metering period. In order to make the clustering algorithm work optimally, all variables were normalized with respect to the smallest and largest (mapped onto a linear scale from 0 to 1) observation.

This step yielded 30 clusters with interesting conclusions. The clusters could be merged based on similarity in their consumption profiles, leading to the consumer groups *Night Owls*, *Late Evening Clusters*, *All Rounders* and *Daily Users*. The interesting conclusion from this is that even without knowledge about the actual consumers, different consumption profiles can be discerned and assumptions about their composition can be made.

The resulting clusters are now used to forecast the energy consumption profile for each cluster. These forecasts can be used by the electricity utilities in order to optimize deployment of different energy sources, in order to minimize their operating costs. The choice to forecast at the cluster level is a trade-off between predicting energy consumption on the national level (too complex) and predicting for every single meter (too much computational effort).

A model is built in which the energy consumption at time t is predicted using consumption data from earlier moments in time ($t-1$ to $t-N$). In order to improve the autoregression, seasonalities on daily and weekly level are first removed. Depending on the cluster, prediction errors going from 1% to 10% were achieved.
Additionally, a neural network (multilayer perceptron) is used to predict the energy consumption time series. Results from this method are not mentioned in the paper.

Initially, the analysis was performed on a laptop with considerable calculation power, comparable to the computer used for this thesis. To compare the regular approach to a big data approach, the study implements a big data analysis using KNIME as well. The difference with the regular approach is that the big data approach uses commercial software that allows distributed computing and hence, faster computation.

2.1.2 Heating energy consumption analytics

Touchie, Binkley and Pressnail (2013) study heating energy consumption in multi-unit residential buildings in Toronto. Their data consists of 40 low, mid and high-rise buildings with consumption figures from monthly electricity and gas bills. The aim of the study is to identify buildings with the highest energy consumption in order to target efforts to decrease consumption efficiently. Further, the influence of several factors (vintage, fenestration, boiler efficiency and ownership) is investigated.

One interesting approach in this study is the energy use analysis: the energy use intensities (EUI, see section 3.3.3) are sorted from high to low. Then, it is assumed that the buildings with the highest EUI can easily achieve the median EUI for the building stock, i.e. by low-cost means such as adjustment controls and replacement of sensors. Further energy savings can be obtained with comprehensive retrofit, albeit at a higher cost, and the lower quartile EUI is used as a savings indicator here.
With only low-cost measures, energy use among the studied buildings can be reduced by 10%. With the high energy savings measures, this number increases to 35% of the current consumption.

The found correlations between building characteristics and energy consumption are less conclusive; the correlation coefficient are lower than anticipated, probably because of the characteristics not being representative for the actual state of the buildings.

2.1.3 Energy consumption in Sweden

Nässén and Holmberg (2005) have made a study about the evolution of residential building efficiency in Sweden between 1975 and 2000. They point out that, though the efficiency increased greatly during the time of the oil crisis of the ’70s, the energy efficiency for the average building has stagnated towards the end of the studied period. They attribute this stagnation to the substitution of heating oil for other energy sources (such as district heating or nuclear power during the ’80s).
Although the scenarios for future energy consumption from 1975 estimated a reduction of Sweden’s energy consumption by more or less 50% in 2000, using newer energy technologies, the consumption appeared to have even increased. Quite contrary to the energy consumption, the emission of CO_2 from energy production have decreased by more than 60% in the studied period, but this reduction seems to be rather because of cleaner energy generation means than because of increased consumption efficiency.

On a different note, Danielski (2012) builds (in part) further upon the previously described paper; he studies the variation in energy (use) intensity for recently constructed residential buildings in Sweden. The buildings are part of the “Stockholm program for environmentally adapted buildings”. This program ran from 1996 to 2005 and aimed at constructing dwellings with an even lower consumption than the building regulations stipulated at the time. However, the studied buildings appear to have a large variation in their consumption per building area.
Danielski explores multiple explanations, some of which are not really relevant to this

study (time interval, size of common areas...). The most interesting observation lies in the dependence of the annual energy use intensity (EUI) on the shape factor, i.e. a measure that indicates how the building's envelope area relates to the floor area. Depending on the shape factor, the EUI in the studied building range from 140 kWh/m^2 (including electricity) to almost the double, while displaying a strong linear correlation.

2.2 Theory

2.2.1 Clustering

According to Webb and Copsey (2011), clustering is the grouping of individuals in a population; the goal is to use these groups to discover patterns in the data. The idea is that individuals in the same group must be as similar as possible, while at the same time, they must be dissimilar from individuals from other groups.

Two cases can be distinguished:

- the data either consists of actual groups with different characteristics, or
- the data has a partly or entirely homogeneous structure.

In the first case, depending on the method and the used parameters (such as number of clusters), the existing groups will be discovered and separated by the clustering algorithm. In the second case, the data will still be divided in groups (often called *partitioned*). In this last case, one must be cautious that the clustering algorithm might suggest a pattern that is not actually present in the data set.

Over the course of years, a vast selection of clustering algorithms has been developed. KNIME implements a few of those, in particular the k-means and hierarchical clustering algorithms. Since Silipo and Winters (2013) use the k-means algorithm in their white paper, this method is adopted in the current study as well.

2.2.1.1 k-means clustering

Following MacKay (2003), the k-means algorithm assigns N data observations in a space of dimension I to k separate clusters. The reason for the *means* in the algorithm's name, is that the clusters are characterized by their I-dimensional mean $\mathbf{m}^{(k)}$ (in which the notation from MacKay (2003) is used). The assignment of data points to a certain cluster k is based on the nearest mean using the euclidean distance, such that

$$d(\mathbf{x}, \mathbf{m}^{(k)}) = \sqrt{\sum_{j=1}^{I} \left(x_j - m_j^{(k)}\right)^2} \tag{2.1}$$

is minimised. Remark that MacKay (2003) uses a slightly different measure of distance, but as long as it is minimal, the k-means clustering works.

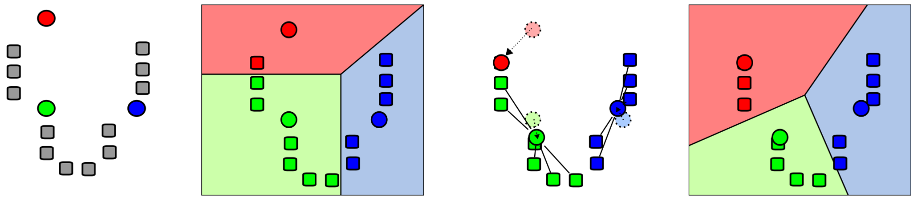
Of course, the cluster means $\mathbf{m}^{(k)}$ are not known from the start. The clustering algorithm is iterative, and the starting condition requires k initial means to be defined as a starting condition. The algorithm is summarised in pseudo code in table 2.1. An illustration of the consecutive steps in the algorithm is provided in figure 2.1.
After the k cluster means have been initialised (fig. 2.1a), all data observations are assigned to the cluster of which the mean is closest to that data point (fig. 2.1b). After all points have been assigned, the cluster means are updated by calculating the average for all points that belong to one cluster (fig. 2.1c). Thereafter, the first step of the algorithm is repeated (fig. 2.1d). As soon as there is no change in the updating of the means, or a predetermined maximal number of iterations has been reached, the algorithm is stopped.

Input	N data points in I dimensions, number of clusters k
Result	N data points assigned to k clusters
	Initialisation
	Choose k cluster means $\mathbf{m}^{(k)}$
	Iteration
while	Cluster means change **or** max iterations not reached
do	1. Assign every data point $\mathbf{x}_j$ ($j = 1 \dots N$) to one of k clusters based on minimal distance $d(\mathbf{x}_j, \mathbf{m}^{(k)})$ (see equation 2.1). 2. Update cluster means $\mathbf{m}^{(k)}$ by calculating the mean for all points $\mathbf{x}_j^{(k)}$ in one cluster: $\mathbf{m}^{(k)} = \frac{\sum_k \mathbf{x}_j^{(k)}}{K}$, where K denotes the number of data points in one cluster. 3. Return to 1.

TABLE 2.1: Description of the k-means clustering algorithm

2.2.1.2 Cluster model complexity

According to Hastie et al. (2009), overfitting (fitting a dataset with too many degrees of freedom) increases the probability that random variations in the data are modelled instead of the actual behaviour that one tries to discover. It is further related to the bias-variance tradeoff in machine learning (Abu-Mostafa, 2012). Bias is the term used



(A) Initialization (B) Assignment (C) Update (D) Re-assignment

FIGURE 2.1: Illustration of the k-means algorithm with 3 clusters. Source: Weston.pace (2007)

to denote the error between the found (or proposed) model and the data set that is used to construct the model from; variance on the other hand refers to the amount of noise (random variation) that is modeled by the algorithm. The combination of bias and variance gives a measure of how well the model predicts the actual behaviour of the phenomenon (in this case district heating energy consumption) outside the studied data set.

A model with low complexity (i.e. a small number of explanatory variables) does not capture much of the random noise in the data set and thus has a low variance. However, the bias is higher because a too simple model in general does not capture much of the actual behaviour either. Hence, the total model has a high error overall. A model with a too high complexity on the other hand may explain all of the variation that is present in the studied data set. However, this will usually also encompass random variations, which is why the performance on data instances outside the testing set of data might be bad again (high variance, low bias). The tradeoff between bias and variance lies in the minimum in the total error that is encountered between these two extreme cases.

Chapter 3

Methodology

Following the outline in the introduction chapter, this chapter provides all information that is needed to understand the steps that were taken in order to produce the results in this thesis.

3.1 Data

This section describes the datasets that are used during this research project.

3.1.1 Energy consumption

The principal data set that is used for the analysis contains energy consumption data for buildings in the district heating network. About 60% of the buildings in Stockholm use DH for heating purposes (Magnusson, 2013).

The energy consumption in MWh is given for 14799 buildings, on an hourly basis for all days of 2012. For each building the corresponding meter ID is supplied. This ID is used to connect building information from other data sets to the energy consumption data.

Clearly, this data set is the largest set that will be encountered in this analysis. With the hourly data for an entire year (2012 being a leap year) for each of the 14799 meters, approximately 130 million rows can be analysed.

3.1.2 Building metadata

A second source of information concerns the buildings. For each meter ID in the energy consumption data, the type of building and building floor area are given. For a number of buildings, the time of construction is given approximately. There are 6 "vintage" classes: before 1925, 1926-1945, 1946-1975, 1976-..., 1976-2005 and after 2006. Depending on the building type, vintage information will or will not be available. Finally, the location of the building is known approximately by means of its zip code.

3.1.3 Weather data

The Swedish Meteorological and Hydrological Institute provides hourly weather data from a vast selection of measuring stations. In this analysis, data for the weather station in Bromma was used to represent the weather in the City of Stockholm. From their database, the hourly values for wind speed, wind direction and temperature have been consulted (Sveriges Meteorologiska och Hydrologiska Institut, 2014).

3.2 Tools

In order to manage the data sources and conduct the analyses, a selection of tools is used. In this section, a brief introduction to each of the tools is given.

Microsoft SQL Server The vast amount of input data is managed in a database. For this task, Microsoft SQL Server was chosen. This software was obtained using the academic Microsoft Dreamspark project, which allows students and academic personnel to use commercial software free of charge.

QGIS In order to analyse spatial phenomena in the energy consumption data, the information is categorised per zip code area and visualised using a Geographical Information System (GIS). The open source program Quantum GIS or QGIS is chosen for this task. It can be freely obtained from the QGIS website. For this project, QGIS version 2.0 is used.

KNIME or Konstanz Information Miner, is an open source data mining tool initially developed at Konstanz University. It can perform extensive data analyses, including reading, combining and processing different data sources, predictive analyses, reporting and visualisation of results. Major advantages of the software are the graphical user interface, which is intuitive and easy to understand, and the vast amount of different analysis tools. KNIME is written in Java and thus runs on all operating systems, but also allows the use of various other languages through a plugin mechanism. The software can be downloaded from the KNIME website. For this project, KNIME version 2.9.2 is used.

http://www.qgis.org/en/site/forusers/download.html
http://www.knime.org/downloads

R is a programming language and software environment that is used for statistical analyses. It can be used as a plugin for KNIME, but as a standalone program as well. R allows to construct graphical analyses of large data sets and allows customization of plots. It is very flexible (adaptability for different categories) to use with different data sets as well, which makes it very suited for big data analytics.

MATLAB is a software environment that allows to perform calculations with matrices and visualise data through a distinct programming language. It was used to perform economic energy savings calculations and compose retrofit scenarios for the studied building set in this thesis. MATLAB is a commercial software package, but KTH offers student licences. The main advantage of using this program is that it allows the user to write functions and to loop over them to investigate the influence of particular variables on the energy savings potential.

3.3 Data processing

To study all hourly consumption values separately would be a very time-consuming task. Therefore, the data is first aggregated (summed and averaged) on various time scales (see below). These aggregation steps are performed in KNIME. Although it's not in the scope of this thesis to present a comprehensive manual of all steps in KNIME, an example of the workflows used in this thesis is shown in figure 3.1.

3.3.1 Time aggregation

The meta-nodes called yearly, monthly, weekly, daily and hourly (see Figure 3.1) execute the aggregation. In these steps, the following values are calculated for each meter ID:

- Annual energy consumption (sum);
- Monthly average consumption;
- Weekly average consumption;
- Daily average consumption; and
- Hourly average consumption

In the discussion and results section, only the annual consumption is used, since all average values are simply scaled versions of this value.

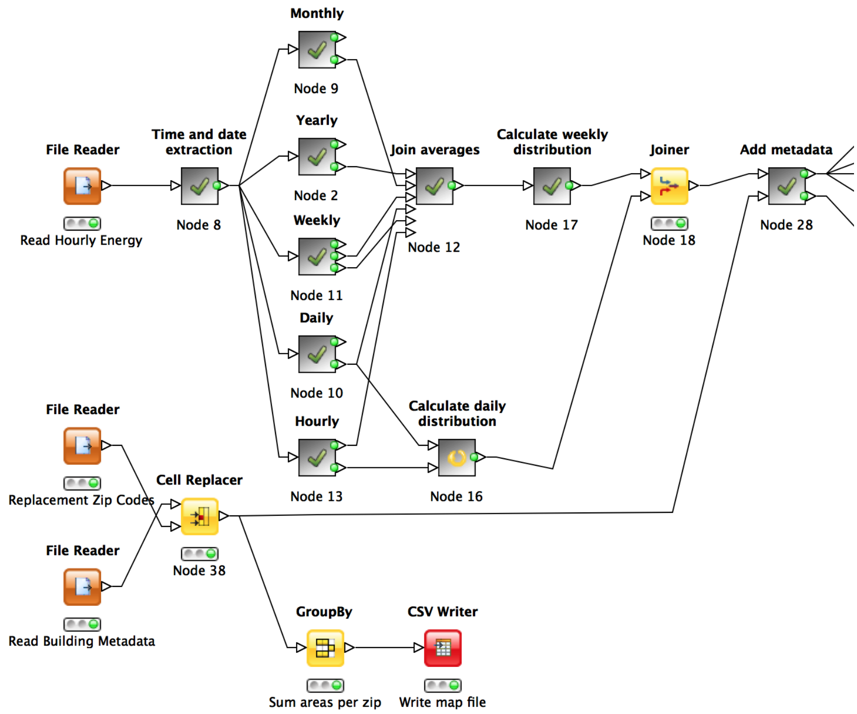


FIGURE 3.1: Reading of energy data and building information in KNIME

3.3.2 Weekly and intra-day distribution

In addition to the time averaging of the consumption on multiple levels, the distribution of consumption over the week and over the day are studied. For the weekly distribution, the division is simple and the proportion of the daily average for each day of the week with respect to the weekly sum is calculated. In addition, the proportions for business days and weekend days are added in separate columns. The intra-day distribution is somewhat less straightforward, since 24 proportions would yield a lot of columns to analyse. Instead, 5 bins of variable duration were defined:

0 - 5 Nightly consumption

6 - 10 Morning

11 - 14 Noon; chosen such to investigate the influence of solar heating during the brightest part of the day.

15 - 18 Afternoon

19 - 23 (Late) evening

3.3.3 From energy to energy use intensity

The buildings in the given data set vary greatly in built area. In order to compare buildings' performances correctly, their energy use is hence divided by the building area. The measure that is obtained in this way, is often called energy use intensity (EUI). Peterson and Crowther (2010) point out that there are many subtleties in the definition of the energy consumption and the building area. For example, the area could be defined as the effective building area, the area that needs to be heated or the area that is occupied by people. Also for the energy consumption, different definitions can be found.

However, the data set does not specify what the energy consumption numbers and areas actually represent, and only one value is given for each of the two variables. Hence, it is assumed that these numbers are consistent within and over the category boundaries and that they can be readily compared.

3.3.4 Box plots

With the knowledge of the EUIs, now the spread of the energy use intensity can be studied. A useful visual tool for this analysis is the box plot.

The box plot summarizes five statistics for a range of observations (McGill et al., 1978), namely *a)* the median, *b)* the lower quartile, *c)* the upper quartile, *d)* the lowest observation within 1.5 IQR from the lower quartile *e)* the highest observation within 1.5 IQR from the upper quartile. The "box" is bounded by the upper and lower quartile values, while the median is indicated with a line inside the box. The "whiskers" extend from the quartile values and end in the extreme values.

The interquartile range (IQR) denotes the difference between the upper and lower quartile observations and can thus be measured from the size of the box. This value is used as a limiting factor for outliers. Although many methods to discern between outliers and regular observations exist, an often used set of rules is the following (Navidi, 2008):

$$\text{Mild outliers if} \quad \begin{cases} \text{Lower Quartile} - 3\ IQR \leq x < \text{Lower Quartile} - 1.5\ IQR \\ \text{Upper Quartile} + 1.5\ IQR < x \leq \text{Upper Quartile} + 1.5\ IQR \end{cases} \tag{3.1}$$

$$\text{Extreme outliers if} \quad \begin{cases} x < \text{Lower Quartile} - 3\ IQR \\ x > \text{Upper Quartile} + 3\ IQR \end{cases} \tag{3.2}$$

In order to compare the variation in spread for different categories, separate box plots are made for each category.

3.3.5 Outlier removal

As seen in equations 3.1 and 3.2 of the previous section, a distinction can be made between mild and extreme outliers. In this study however, this difference is not used; the outliers serve the purpose of identifying wrong data more easily.

In usual data exploration, extreme outliers can be removed to make the analysis more accurate. In the topic of DH, this does not seem to be a good strategy. It can be expected that some buildings may have a remarkably high energy consumption, while they should not be seen as outliers. To keep the analysis as comprehensive and complete as possible, only buildings with a seemingly unrealistic energy use are removed.

In the list of outliers, the buildings with unrealistically high annual EUIs were studied separately. If an anomalous consumption profile (e.g., a constant consumption for several weeks, negative values, very high consumption peaks for one hour. . .) was discovered, this building's data was removed from the working database.

In some specific cases, the energy consumption profile was not anomalous. Instead, the high EUI could be explained by an unusually small building area. In order to cope with this problem, buildings with an area smaller than or equal 10 m^2 were removed from the data set as well.

Some building categories ("Transit", T-bana stations, thermal plants. . .) only contained a small number of buildings. These special categories were not included in the general classes in order to be studied separately (for more information about these classes, see paragraph 3.6.1). No outliers were removed from these categories.

3.4 Making maps

A large part of the energy analysis in this thesis is done by making maps of the available information. The steps to be followed to complete these task are always the same; therefore, the mapping procedure will be explained only once in this separate section. In addition, some of the problems that were encountered and their solutions will be explained.

3.4.1 Choropleth

A choropleth is a map that visualizes the value of a certain variable in a confined area using a range of colours (Lloyd, 2010). In the case of this research project, the variable is in most cases the energy consumption or a related value. The confined areas are the zones with the same zip code. In this case, it was an evident choice, since the information on the buildings' location is available in the form of its zip code. However, even if information on a finer scale (e.g., the address) would be available, the zip code areas would still be a better choice for map readability.

3.4.2 Construction of a choropleth

As stated in the tools section (3.2), QGIS was chosen for making maps. The base for the maps is always a shapefile (`.shp`-extension) that contains the contours of the zip code areas in Stockholms Län. Attached to the zip code areas is a label with the actual zip code. In order to assign a certain variable to these areas, a join statement in QGIS is used. As in KNIME and SQL, the entries of two data tables are combined (*joined*) based on the equality of the value for one column in both tables. In this case, the equality of the zip code in both tables is checked. Once the variable that has to be mapped is added to the shapefile attributes, this attribute must be converted from a string to a number (in most cases decimal), or otherwise QGIS cannot process the values.

After this conversion, the appearance of the layer must be changed. By default in QGIS, all zip code areas have the same colour; this setting must be changed to a gradual colouring, which implies that the colour of the area depends on the value of the variable that is defined by the user. Now, a number of intervals is defined, and each interval receives a different colour. Based on the classification method, the intervals will have different boundaries. For the maps in this study, the Jenks classification method was chosen. This method defines the breaks between the different classes such that the variation of data within the separate classes is the smallest, while making sure that the classes are far enough apart (Jenks and Caspall, 1971).

3.4.3 Missing zip codes

In the data set that was provided, only a part of the zip codes in Stockholms Län was used. On the other hand, the set also contained about a hundred zip codes that did not occur in the shapefile in QGIS, which were identified from the error log after the join statement. Using the web site of the Swedish *Posten*, the missing zip codes could be traced; either:

- The zip code did not exist,
- the zip code referred to a postal box number, or

`http://www.posten.se/sv/Kundservice/Sidor/Sok-postnummer-resultat.aspx`

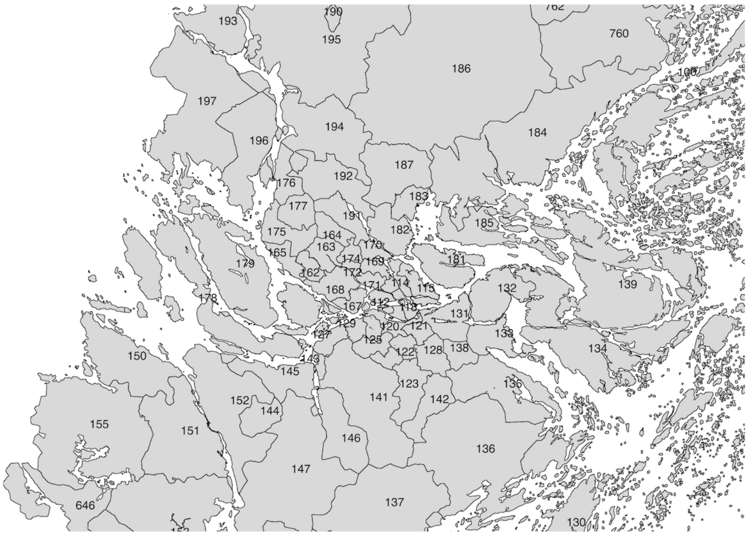


FIGURE 3.2: Map of the areas with the same zip codes, based on the first three digits

- the zip code did exist, but was not included in the map yet.

In case the zip code did not exist, the zip code was replaced by 0 in the KNIME analysis. The same was true for the box zip codes. However, a small number of zip codes with no physical location was used by municipalities surrounding Stockholm (e.g., Lidingö, Upplands Väsby…). In this case, the consumption data under these zip codes were assigned to a real zip code near the center of that community.

In case the zip code did exist according to Posten, but was not included in the map, the base shape file was altered. In some cases, the shape of the missing zip code area could be viewed in Google Maps and hence drawn in QGIS. Sometimes even Google Maps did not show the zip code, and the area had to be approximated using the addresses listed for that zip code by Posten. The drawing of the new areas was performed in QGIS by splitting existing zip code areas to form a new area.
It was observed that most missing zip code areas that were added according to these steps were recently developed or even to-be-developed parts of the city, such as Norra Djurgårdsstaden (Stockholm Royal Seaport) and Hagastaden.

3.4.4 Three digit zip code maps

Because of the irregular size of Stockholm's zip code areas, it will prove useful for the energy maps in this thesis to study the data in larger, more regular areas. Because of the way the zip code system in Sweden works, the areas which share the same three digits of their zip code are suitable for this purpose. An example of these areas is shown in Figure 3.2.

The zip code shape file on three digit level was not freely available. Therefore, the base shape file had to be adapted again in QGIS. In the attribute table of the base shape file a simple string operator was applied to the postal numbers. This yielded the three first digits of the postal number in a separate column. Now, all areas with the same 3-digit number could be selected and merged (using the "Digitizing" tool bar).

Choropleth maps can be constructed from this adapted shapefile as well. The only difference is that when data needs to be added to the shapefile attributes, the 3-digit zip code must be used as the joining criterion instead of the full postal number.

3.5 Clustering

This section describes the application of the theory in section 2.2.1 in order to derive patterns from the data set. Since the different classes have been shown to behave quite differently, this information is already used before applying the k-means algorithm, in the sense that the clustering is performed for each class separately.

The next step is to decide which dimensions will be used for the clustering algorithm. The goal is to group buildings by the amount of energy that they use, their use intensity and the moment of use. Therefore, the following parameters are selected:

- Annual EUI,
- annual total consumption,
- proportion of use during different times of the day and
- proportion of use on weekdays/weekend days.

Caution is needed here, however. The goal of k-means clustering is to find the right (number of) clusters, if present in the data set. From a data set with as much information as the current one, it is hard to say beforehand which dimensions decide on the difference between clusters. Chances are that the proposed set is too small, but the danger is even greater when the set is too large. In this case, variables that are correlated might affect the result of the clustering algorithm. Since the variables that represent fractions always have a sum of 100%, one of the variables can be left out, since it is a linear combination of the remaining variables. Using this method, the later evening proportion (19-23h) and weekend (Sat-Sun) variables are deleted from the list.
In the same way, it can be found that it is unnecessary to include either the size of the buildings, since it can be derived from the EUI and annual consumption, or average values on different time scales – because they are just scaled versions of annual averages.

While the k-means algorithm is very proficient at finding clusters in a data set, it is not able to determine the value of k that is needed (cf. section 2.2.1.2). A number of rules of thumb, indicators and other techniques to find k have been discussed (see for example Tibshirani et al. (2001)). While the author has experimented with a few of these measures, the amount of data seems so big that an inconveniently large number of clusters would be needed, in addition to a lot of calculation time. Therefore, the number of clusters has been heuristically limited to 20. It is believed that this number is large enough to discern the several types of consumers that might be present in the provided data, while at the same time it is small enough to distinguish between a number of clusters without too much effort.

Since the k-means algorithm is distance based, it does not perform well on data sets that have large differences in spread for the investigated explanatory variables. The typical example is that of two elongated clusters (MacKay, 2003, pg. 20), in which the k-means algorithm fails to distinguish the actual clusters. Since this particular data set inherently displays a large spread in annual consumption, and on the other hand smaller spreads for the proportional variables, it is necessary to normalise these variables. In this case, the simplest normalisation method is used: the smallest observation is mapped onto 0, the largest onto 1, and all intermediate observations are scaled linearly between the two extremes. This normalisation step was abridged from Silipo and Winters (2013).

A maximum number of 400 iterations was chosen. However, in most of the cases the algorithm reached convergence long before this limit.

3.6 Maximum energy savings potential analysis

Similar to the method used by Touchie et al. (2013), this study also imposes different EUI limits to the buildings in order to investigate the effect in terms of energy savings. Because of the high variety in the building stock, the analysis was conducted separately for 6 building classes.

3.6.1 Building classes

In order to maintain a simple overview of the buildings, while at the same time preserving the distinction between the various categories, six classes were formed to replace the 30 different categories in the data. The definition of these classes is explained in table 3.1. Remark that a number of categories (e.g., thermal plants, street heating and subway stations) were not included in these classes because of particular behaviour. These categories will have to be studied separately.

Class	**Contained categories**
Residential	• Single-family & detached house, • Multi-family house with additional function, • Apartment in detached multi-family house, • Detached multi-family house and • Multi-family house in building block.
Commercial	• Office & Shop buildings, • Office building with additional function, • Hotel and • Warehouse with office.
Public	• School, • Sports facility, • Community & meeting center and • Church.
Health & Care	• Hospital or nursing home, • Service flats and • Daycare & recreation center.
Industrial	• Printing industry, • Workshops, • Garage, • Vehicle assembly and • Other (industrial) facilities.
Other	• Only warm water, • Only heating, • T-bana (subway) station, • District heating plant, charged and • District heating plant, not charged.

TABLE 3.1: Definition of the building classes

	Electric heating	Other heating
Houses	55	90
Other buildings	55	80

TABLE 3.2: EUI regulations for climate zone III as stated by Boverket (2013) (values in $\frac{kWh}{m^2 \cdot a}$)

3.6.2 Energy savings scenarios

In order to assess the savings potential for the building stock in the City of Stockholm, the consequences of different energy use limits can be studied. Since the total consumption for a building depends strongly on the heated volume (or by proxy in the studied data, the building surface) and the buildings vary greatly in size, it is appropriate to limit the EUI instead of the total consumption.

To assess the total amount of energy that is saved annually over the building stock per class, the EUIs need to be multiplied by the building areas again:

$$E_{saved,tot} = \sum_{i=1}^{n} (EUI_{actual,i} - EUI_{lim}) \cdot A_i, \tag{3.3}$$

where i denotes the buildings for one building class and A_i is its built area.

Various EUI limits can be imposed for different building classes. As described by Touchie et al. (2013), these limits can be derived from the data set that is currently studied. In this case the median EUI for each building class can be applied as a *low-cost* savings limit; the lower quartile EUI value can be used as a *comprehensive retrofit* savings indicator. In this thesis, only the median value is used as a savings indicator.

Instead of deriving the other EUI limits from the upper and lower quartile EUIs, two additional saving scenarios are based on Swedish energy regulations. Boverket (2013, chap. 9) has set limits to the EUI value for new buildings of different categories, depending on the type of heating installation and the climate zone. Sweden is divided in 3 climate zones because of the differing climate types in the North, Center and South of the country (see figure 3.3). Stockholm lies in climate zone III. The according EUI limits are listed in table 3.2.
Notice that in the actual building regulations, a correction is made for buildings that have a larger ventilation system. Since no information about the type of ventilation is available in this data set, the lower EUI limits can be seen as a extreme case of the potential savings.

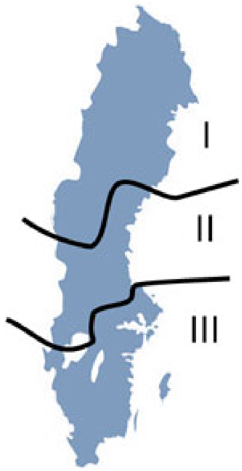


FIGURE 3.3: Illustration of climate zones in Sweden. (Boverket, 2011)

3.7 Retrofitting plan

In Stockholm's action plan for environment and energy, Lönngren (2012) sets out the path to be taken in different energy consumption groups (buildings, transport etc.) in order to reduce the total energy consumption and emitted GHGs for the city. In the current thesis, the suggestions from this report are applied to the studied building set, and it is investigated what the economic picture would be if the goals that were set for 2015 are to be reached.

3.7.1 Goals and method

The action plan (Lönngren, 2012) mentions a number of energy and climate goals set for 2015. Among others, the City specifies that all property in the city should reduce their combined energy consumption by 5% w.r.t. 2012. This study tries to identify which measures should be taken in the given data set in order to reach this goal. It must be said that this study only uses heating data, while an important share of the total consumption consists of electricity.

The next paragraph defines a number of (hypothetical) savings measures. In addition to the energy perspective, also the cost and profitability of these savings measures are estimated. In order to assess the efficiency of investment in one building, the non-discounted return on investment (ROI) is used:

$$ROI = \frac{t_{inv} \cdot R - I}{I}, \tag{3.4}$$

with t_{inv} the investment horizon, R the annual returns (in this case the cost of saved energy) and I the investment. The reason that rates of intrest are not included in this measure, is because these rates fluctuate highly for this type of investment, and the measure only serves the purpose of comparing the investment among the studied buildings.

Once the new EUI is assigned and the annual savings and ROI are calculated, the buildings are sorted by descending ROI. The energy savings are summed and the first x buildings that together achieve the set savings goal are presented as a *retrofitting plan*.

3.7.2 Measures and costs

In this methodology, three types of savings are considered, in accordance with the city's action plan. *1*) Small savings in energy consumption can be achieved by updating the heating control system (new thermostats). This can reduce the initial consumption with 5% and has a small cost per square meter (30-50 kr/m^2). *2*) Somewhat larger savings include the replacement of thermostats from *1)* and replacement of the ventilation system. Ventilation losses constitute 33% of heat consumption, and by installing new systems with a heat exchanger (heat recovery ventilation system or HRV), 85% of these losses can be saved, i.e. 28%. Combined with the improved thermostats, this yields 33% savings at a cost between 500 and 1000 kr/m^2. *3*) The largest savings are achieved by retrofitting the building shell. Heat losses amount to 84% of the heating energy consumption, and the climate action plan estimates that about 70% of these losses can be saved by a shell retrofit, so 59%. A complete retrofit is more expensive: 2000-3000 kr/m^2. However, the author believes that buildings with very high EUI values could perform much better after a shell renovation; therefore, the scenario is implemented with a fixed EUI after retrofitting, in this case 60 kWh/m^2.

Which savings measure is applied depends on the current consumption of the building. Applying the regulations from Boverket (2013) again, it is suggested that the new energy use intensity of retrofitted buildings should lie between 55 and 80 kWh/m^2. Then the first measure should be applied in buildings with an EUI between 60 and 80 (yielding new EUIs between 57 and 76), the second measure between 80 and 120 (after savings 54 and 80) and the shell renovation above 120 kWh/m^2. This method is illustrated in figure 3.4

In addition to these savings, Sjögren (2007) argues that by using newer technologies (e.g., low-flow water taps), an energy saving of around 5 kWh/m^2 can be realised. In this study, we assume that buildings that are built after 2006 already use these technologies. In all other buildings, this additional saving is subtracted from the annual EUI. The cost thereof is assumed to be insignificant because of the low pay-back time (Sjögren, 2007).

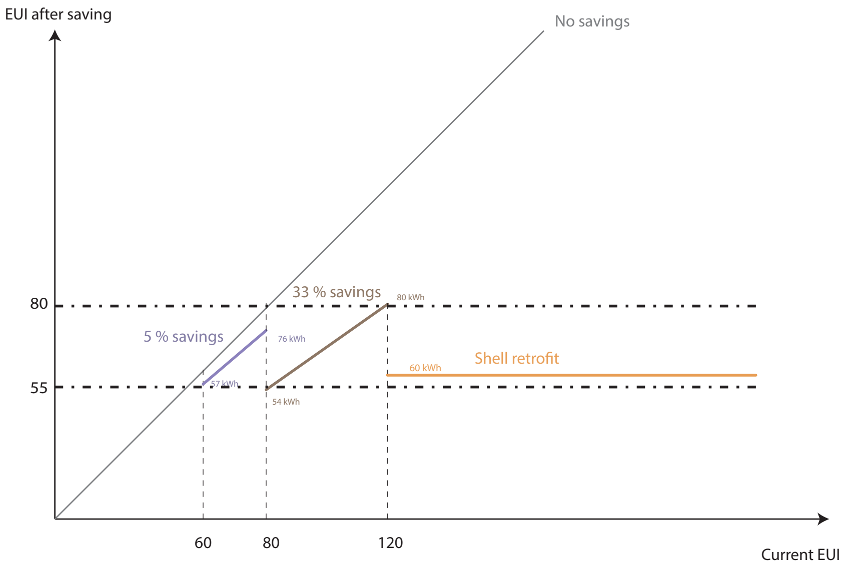


FIGURE 3.4: Illustration of EUI limits for retrofitting scenarios

On the other side, the energy savings can be expressed using the cost per unit of energy for the district heating system. Although the cost depends on the type of contract and customer, a lower limit of 900 kr/MWh can be derived from these estimates (Fortum, 2014).

3.8 Weather influence analysis

It is anticipated that in a city, the influence of weather on buildings' energy consumption may depend on their location. Buildings in the city centre might be more shielded from wind and cold because of the higher building density, whereas in the suburbs, the opposite may be true. This section describes how the analysis for this effect was set up.

3.8.1 Assumptions

3.8.1.1 Linearity of weather influence

It is assumed that the dependence of the energy consumption on temperature and wind speed can be described as a linear relation. This is only partially correct.

On the one hand, it is true that the outflow of heat from a building is proportional with the difference between the inside and outside temperature, as illustrated by the one dimensional heat flux through an infinite plate,

$$Q = kA\frac{T_{out} - T_{in}}{L} \quad \text{or} \tag{3.5}$$

$$Q = hA\Delta T \quad \text{with } h = \frac{k}{L}, \tag{3.6}$$

with Q the exchanged heat, A the area through which the heat is flowing, k the thermal conductivity and L the thickness of the wall. h is the heat transfer (Lienhard IV and Lienhard V, 2008).

On the other hand, the heat exchange caused by the wind, which can be modelled as convection, does not have a simple linear relationship with the wind speed. Again, the heat exchange depends on the temperature difference, but now the heat transfer coefficient h is determined by a number of parameters in the convection process. One of these parameters is the fluid (in this case the stream of air or wind) and it may be assumed that h varies proportionally with a power of of the air velocity.

Because of the small fluctuations in wind speed and the unavailability of exact wind speed data at the location of every measuring point, we will further assume that a linear relation is sufficient to discern differences in the weather dependence, if present.

3.8.1.2 Energy normalisation

Because of the large variety in building sizes and hence total annual consumption (see paragraph 3.3.3), again the EUI (summed per day) was used as a measure for energy. But still, then the consumption still depends on other (hidden) factors than weather alone, e.g. income, number of inhabitants, personal preference for a higher or lower indoor temperature…

In order to cope with these external factors, the daily EUI values were made dimensionless by normalizing them to the annual EUI, as shown in equation 3.7.

$$EUI_i^* = \frac{EUI_i}{\sum_{j=1}^{366} EUI_j} \tag{3.7}$$

3.8.2 Regression analysis

In order to compare the influence of weather aspects on the energy consumption in various (types of) buildings, a linear regression model is built. In this model, the dependent variable is the dimensionless EUI^*, and the independent variables are outside temperature T and wind speed v. The model equation can then be written as

$$EUI^* = \beta_0 + \beta_T T + \beta_v v + e, \tag{3.8}$$

with β_0 the intercept, β_T and β_v the temperature and wind speed coefficients and e the prediction error. Using a least squares algorithm, the parameters β are estimated such that the prediction error is minimised overall. When the coefficient for wind or temperature is relatively higher in absolute value, it could be inferred that the influence of the respective variable is higher.

In order to compare the influence according to location, consumption and building type, the models are partitioned according to these criteria. In a first attempt to discern local variations, the regression models are divided per 3-digit zip code. This method could be seen as a simplified version of spatial regression (Lloyd, 2010; Wheeler and Paéz, 2010), in which the weight for observations within the same 3-digit zip code area is 1, and 0 for all the others. In this way, separate coefficients are obtained for all zip code areas.
Further, the buildings are classified according to their annual EUI, viz. in four classes:

- above 200 kWh/m^2,
- between 160 and 200 kWh/m^2,
- between 100 and 160 kWh/m^2 and
- below 100 kWh/m^2.

In this way, although the largest fluctuation in annual EUI has been cancelled by using the dimensionless quantity, differences across buildings that are better or worse consumers can still be observed.
As a final distinction, also models for different building types can be constructed. In the stock of residential buildings (which are the main subject of this regression analysis), there are five classes (see table 3.1).

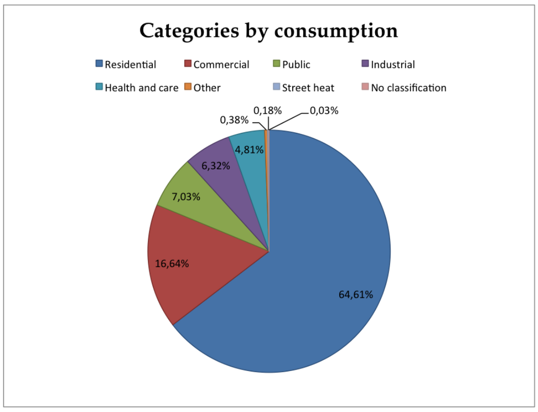
Chapter 4

Results

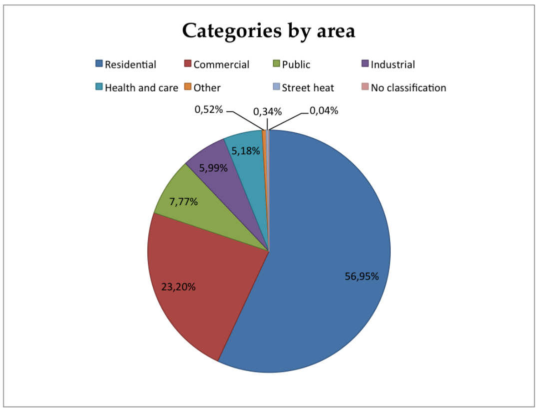
4.1 Descriptive statistics

This section tries to provide a description of the averages and ranges that are encountered for the different building categories. The results are grouped per building class (see section 3.6.1) and where possible, information about vintage type of the buildings is provided as well.

Before putting emphasis on the separate categories in each class, first an overview of all categories is given in the form of pie charts (figure 4.1). Figure 4.1a calculates the proportion of each class based on annual energy consumption, while figure 4.1b uses the building area. Both charts show more or less the same distribution, except for the residential class, which has a higher consumption percentage compared to its area, and the commercial class, where the opposite is true. These differences are present, but less outspoken in the remaining classes and categories.



(A) By total annual consumption



(B) By total building area

FIGURE 4.1: Pie charts of building classes

In addition, the boxplots for each separate building category can be found in figure 4.2 and 4.3. Remark that the scales of the EUI and area axes have been confined and that some of the extreme outliers are not shown because of this.

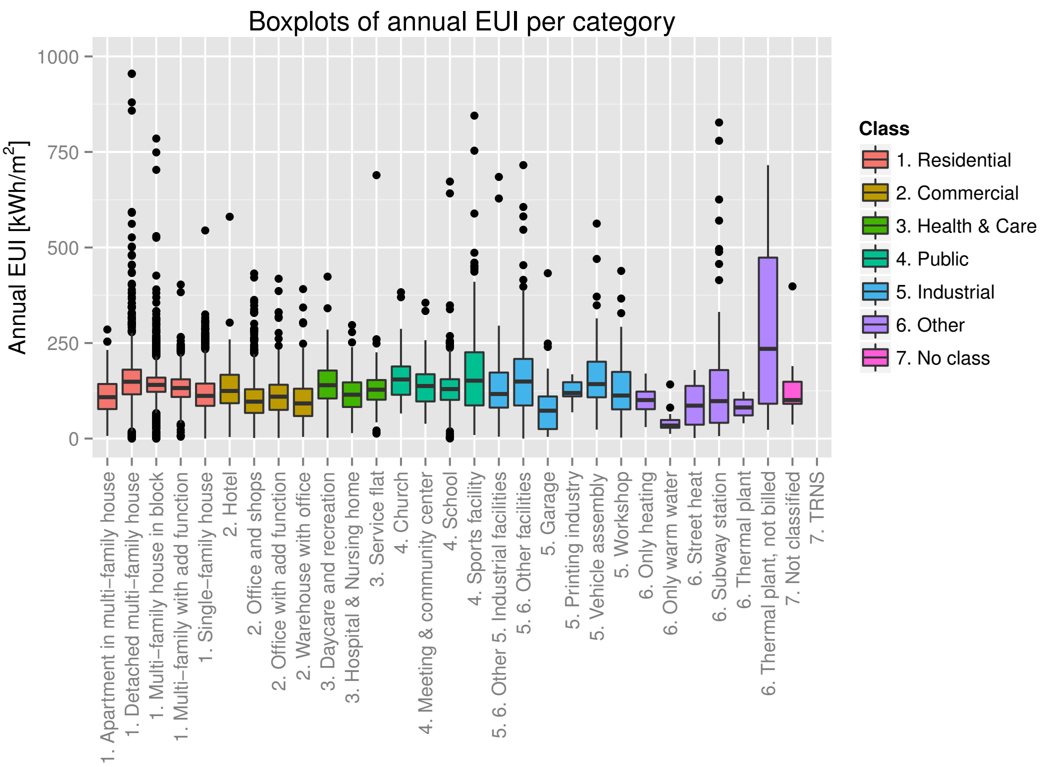


FIGURE 4.2: Boxplot of annual EUI per building category

The following paragraphs give a brief description of the descriptive statistics for the various building categories. For information about the spread of EUI values per class, the reader is referred to figure 4.2. Appendix A presents comprehensive summary statistics for each building category for the total annual consumption, EUI and building area.

4.1.1 Residential buildings

Table A.1 provides a statistical summary of the different categories within the residential building class. The mean EUIs can be observed to be higher for the larger buildings, while single apartments and smaller houses have lower consumption intensities. The standard deviations on the different classes have the same order of magnitude, although the meter counts are quite different.
As anticipated in section 3.3.3 the standard deviation and range for the total consumption numbers are much larger than for the EUI values.

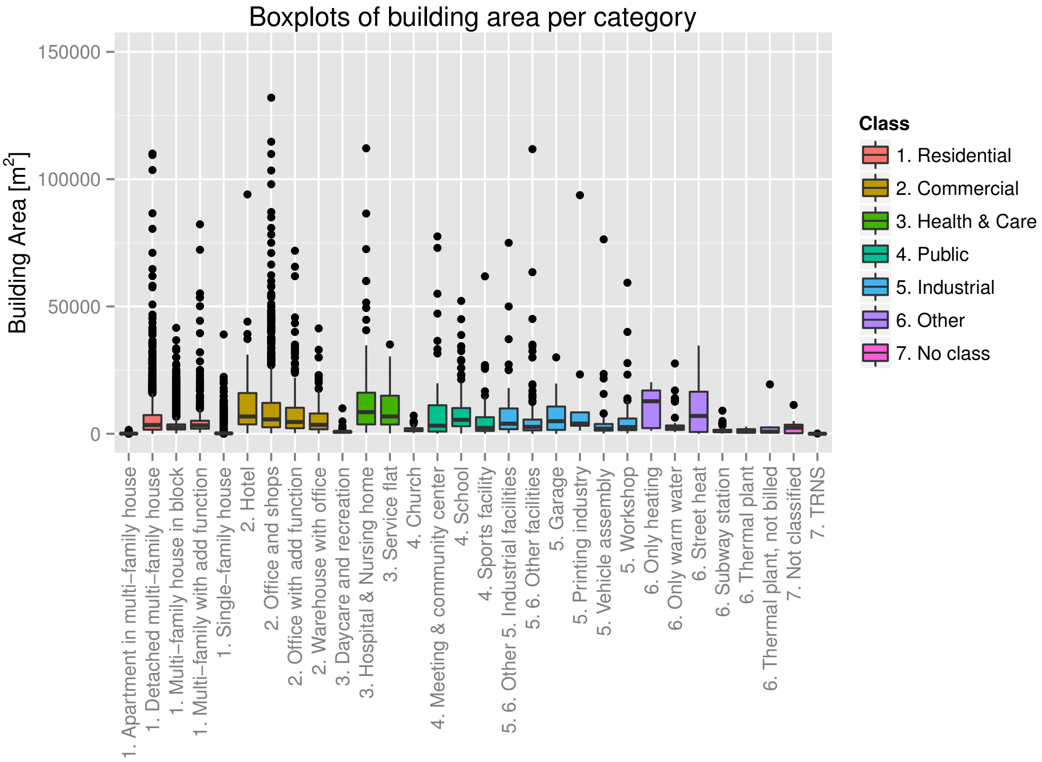
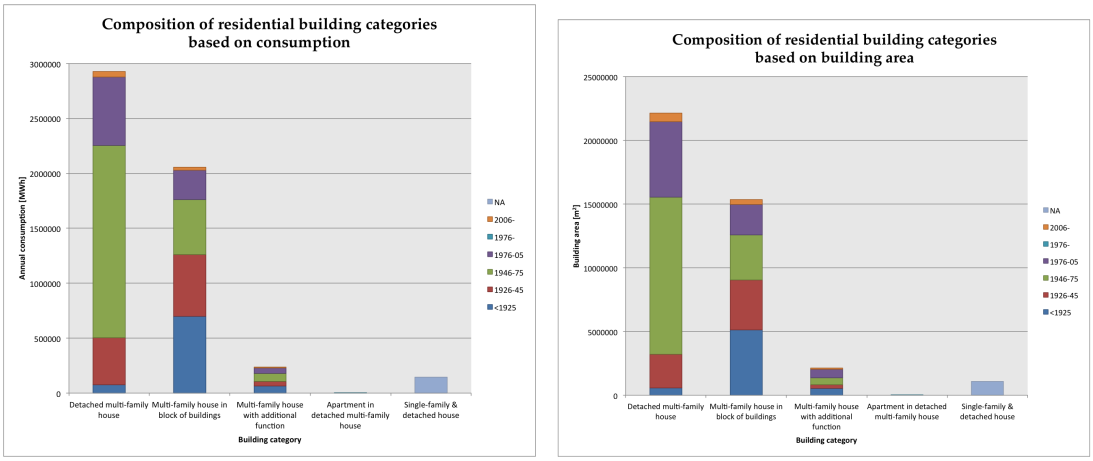


FIGURE 4.3: Boxplot of building area per building category

Both for residential and commercial meters, information about the age of the buildings is available in the form of the vintage class. Figure 4.4a and 4.4b explore the composition of the residential building class based on category and vintage. While the former shows the composition in terms of annual consumption, the latter is based on the floor area of the buildings.

Both show a more or less equal distribution, with most of the buildings being multi-family houses, either detached or in a block of buildings. In the first group, the majority of the houses was built between 1946 and 1975, while the division is more equal in the second class. The remaining three classes are far less in number.
An interesting detail about the data set is that there is no vintage information only for the single-family houses, while information is available for all other residential categories. Further, the single apartment category is the only one with the label *1976-* instead of *1976-2005* and as a matter of fact, all buildings in this category are from this specific time period.

In order to make the subtle differences between the compositions based both on area and energy consumption clear, a third graph can be plotted. In figure 4.5, the total EUI for the building categories per vintage period are indicated. The EUI here is calculated as the sum of the annual consumptions for all buildings from the same building period



(A) Residential annual consumption (B) Residential building area

FIGURE 4.4: Composition of residential buildings in terms of categories and vintage

and category, divided by the sum of their areas. This method enlarges the differences between the compositions in figures 4.4a and 4.4b.

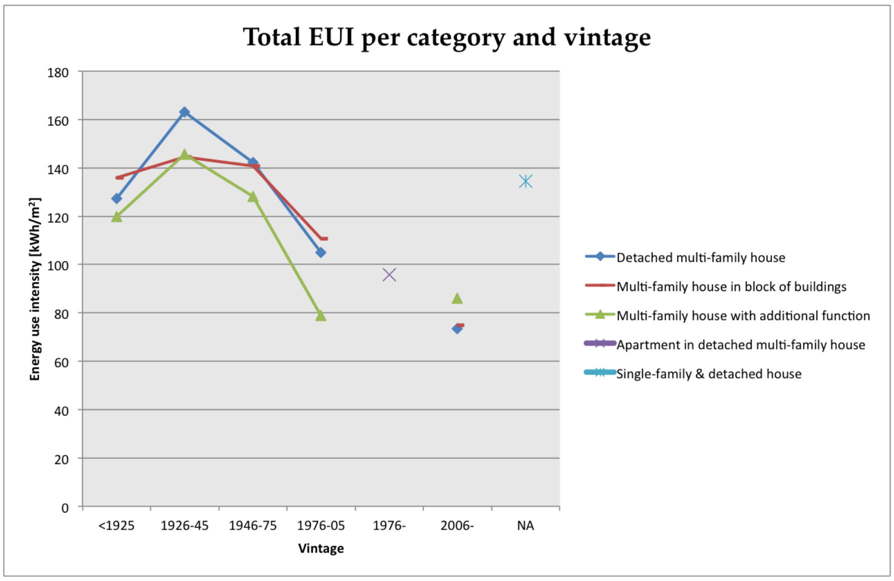


FIGURE 4.5: Total EUI per category and vintage period (expressed in kWh/m^2) for residential buildings

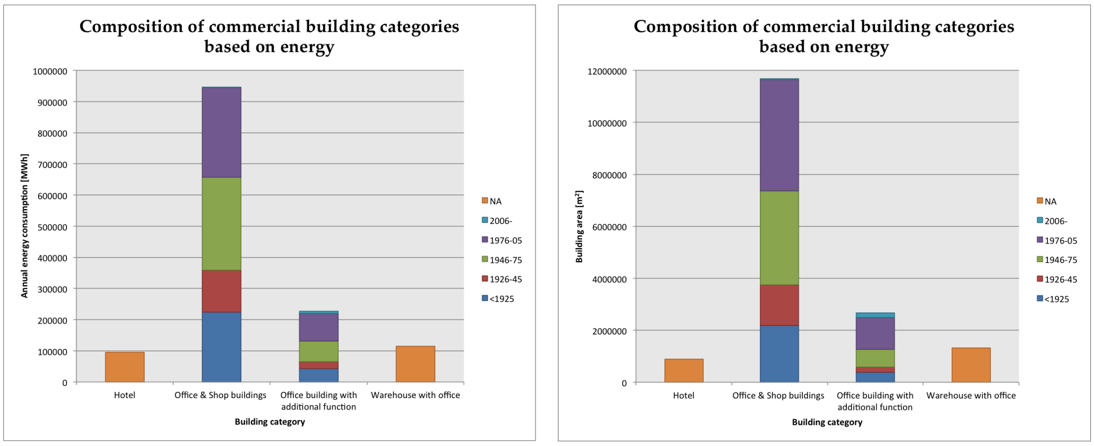
As pointed out by the differing EUI values, the compositions based on energy and area are not directly scalable. An interesting observation is that similar temporary evolutions can be observed for the three largest categories: maximum intensity for the buildings from *1926-45*, then decreasing to lower intensities until the most recent buildings. One exception is the category with additional functions, which remains at the same EUI from 1976 onwards. This category has the lowest EUIs overall, while the buildings in block have a lower peak EUI, but the highest values for *-1925* and *1976-05*.
The apartment category appears separately in the *1976-* vintage period, and performs

according to the EUIs of the other building categories in the same time period. Referring back to the paper by Nässén and Holmberg (2005), this illustrates the stagnation of energy efficiency in the last quarter of the 20th century. The single family houses have a rather high EUI, but it must be remarked that this is an average value for all building periods, which remain unknown.

4.1.2 Commercial buildings

The averages, standard deviation and extreme values for the different categories of commercial values are gathered in table A.2. It is immediately clear that the number of buildings of this type is smaller than the amount of residential buildings. Again, the high variance in total consumption as opposed to a low variance in EUI is apparent. Remark the lower mean EUIs when compared to the residential use intensities. At the same time, the commercial buildings spread a larger floor area than their residential counterparts.

Using the same method as for the residential buildings, the composition of the commercial building class is shown based on energy in figure 4.6a and based on building area in 4.6b. Quite similar compositions are shown again, with a high amount of office and shop buildings, and a smaller amount of these with an additional function. The hotel and warehouse categories are less represented, and have no vintage information moreover. The composition of the two largest categories seem similar in proportion, but the additional function category has more buildings from 2006 onwards pro rata.



(A) Commercial annual consumption (B) Commercial building area

FIGURE 4.6: Composition of commercial buildings in terms of categories and vintage

With the previously described technique, the overall energy use intensities per vintage and category can be calculated. The resultant numbers are showed as a graph in figure 4.7. The two office buildings categories display a very interesting evolution: the oldest buildings start out at an overall EUI around 100-110 kWh/m^2, but quickly decrease to only 40 kWh/m^2 in the newest buildings. These values are all lower than the EUIs for residential buildings. In this case, there is no direct evidence of the efficiency stagnation that Nässén and Holmberg (2005) point out.

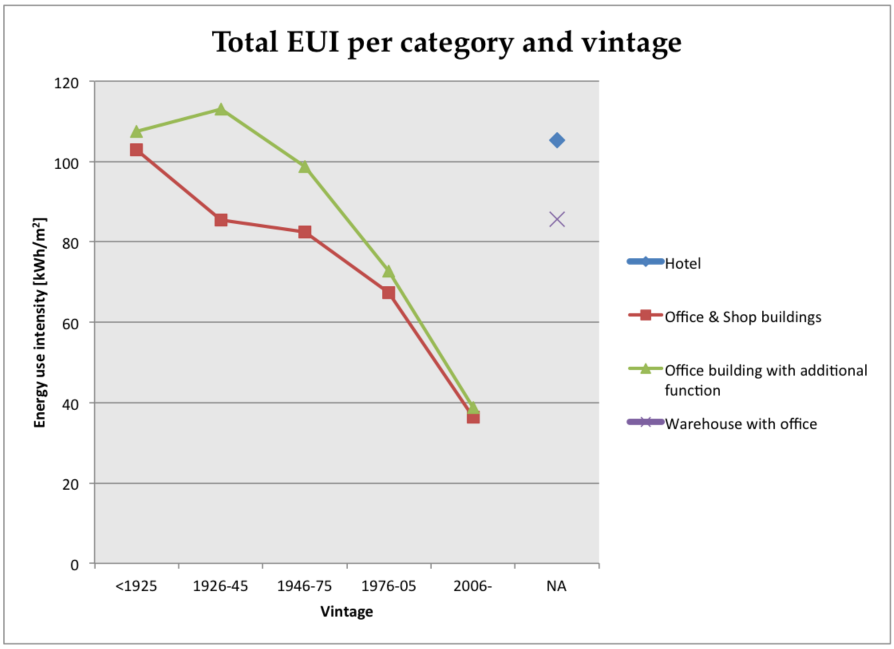


FIGURE 4.7: Total EUI per category and vintage period (expressed in kWh/m^2) for commercial buildings

4.1.3 Health & care buildings

Table A.3 summarizes the descriptive statistics of buildings for the health and care class. The energy use intensities for these categories are situated in the same range as the residential meters. The high annual consumption for hospitals and rather small consumption for daycare buildings follow from the annual energy consumption means.

Furthermore, the respective shares of building categories for this class are presented in figure 4.8.

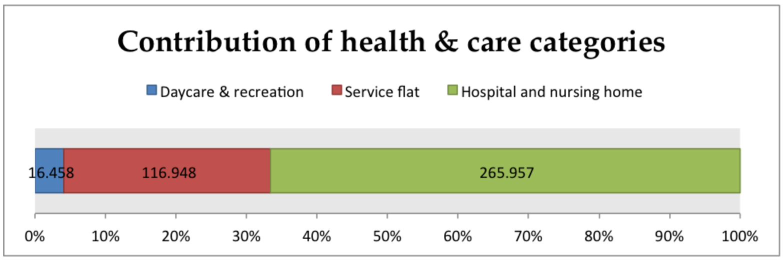


FIGURE 4.8: Shares of the various health care building categories by annual energy consumption [MWh]

4.1.4 Public buildings

The descriptive statistics for the buildings in the public class can be found in table A.4. The higher EUI means are remarkable; furthermore, the spread on the consumption intensities seems to be higher than in the previously studied classes, pointing to a higher heterogeneity within the public building class.

Again, the shares of the constituting categories are presented in figure 4.9

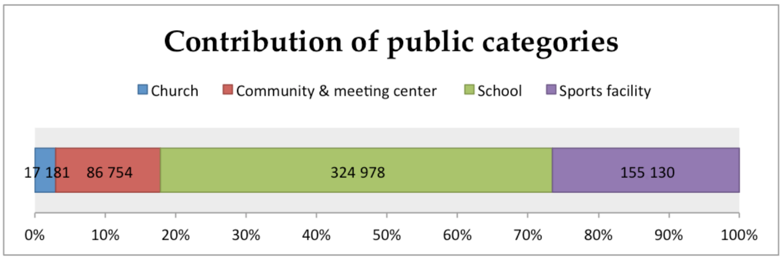


FIGURE 4.9: Shares of public building categories by annual energy consumption [MWh]

4.1.5 Industrial buildings

Table A.5 shows the descriptive statistics for industrial buildings. The heterogeneity of the industrial buildings is reflected in the wide range of mean EUIs per category, Moreover, the standard deviation on the EUI is much higher when compared to the previous results. The respective shares in energy consumption are visualised in figure 4.10.

Remark that there is still one building with an area of 1 m^2. This meter ID has slipped through the outlier removal mechanism, most probably because of a still acceptably low EUI value. Instead the category next to that, with other industrial facilities, appears to have the highest maximum EUI with a surprising value of 12 000 kWh/m^2.

A second comment on the industrial buildings is that the high variation in both EUI and total consumption may be explained by the fact that heat from the DH network is possibly used for other purposes than heating alone. In industrial buildings, steam from the heating system can be used as an input for production processes.

4.1.6 Other buildings

Finally, table A.6 summarizes the descriptive statistics for the remaining building categories. The few buildings with either only warm water, or only heating have rather low (maximum) EUI values. The T-bana (subway) stations have normal consumption figures, compared to the previous cases, although the standard deviation is rather high.

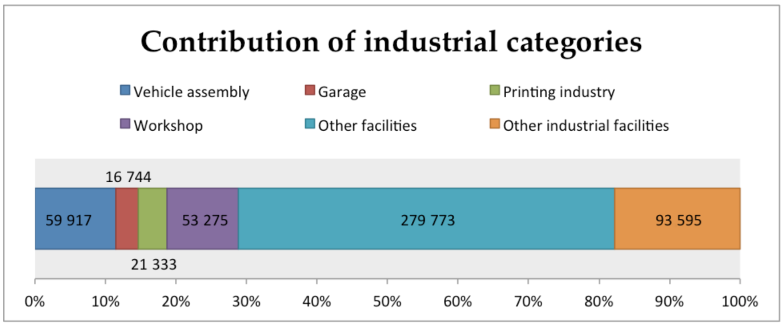


FIGURE 4.10: Shares of industrial building categories by annual energy consumption [MWh]

The 8 district heating plants that have not been identified as outliers display on the one hand rather low EUIs (charged), on the other hand quite high values (not charged).

Street heat is a special class of heating, which shows a remarkably low EUI. The maximal value as well as the standard deviation fall in an equally low category. The buildings with no classification are only incorporated for completeness and will not be analysed further.
The shares in annual energy consumption for the categories within the "other" class, except street heat and non-classified buildings, are shown in figure 4.11

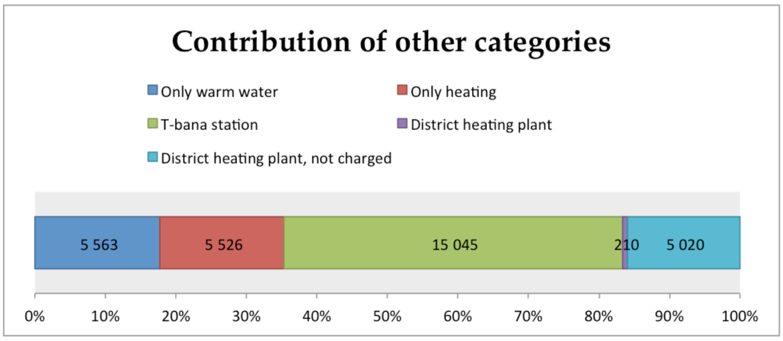


FIGURE 4.11: Shares of other building categories by annual energy consumption [MWh]

4.1.7 *Key findings*

- Residential buildings account for the largest consumption share, followed by commercial buildings
- Buildings from before 1975 are predominant, with 1946-1975 as the largest age group for residential buildings and 1976-2005 for commercial buildings.
- There is a downward time evolution for the EUI in those two classes, but there is a stagnation for the most recent time period
- Very high EUI values occur in the industrial buildings class

4.2 Time-dependence of energy consumption

Improving energy efficiency and reducing GHG emission is not just a question of reducing energy consumption in absolute numbers; one has to keep in mind that by decreasing the peak consumption, the use of polluting peak capacity generators can be made unnecessary. Therefore, it is important to study the instantaneous power demand for district heating as well.
In this case, the power is not known, but the hourly consumption is. The power can thus be approximated by dividing the hourly energy consumption by the time interval:

$$P = \frac{dE}{dt} \approx \frac{\Delta E}{\Delta t}. \tag{4.1}$$

In other words, the hourly consumption in MWh equals the average power demand in MW for that time interval.

4.2.1 Annual

First, a general image of the daily energy consumption throughout 2012 is formed. Figure 4.12 shows the sum of all buildings' consumption, grouped per day. Apparently, consumption is highest around the 30th day of the year, i.e. end January/begin February. As expected, consumption is lowest during summer, but there is a remarkable peak between day 150 and 160.

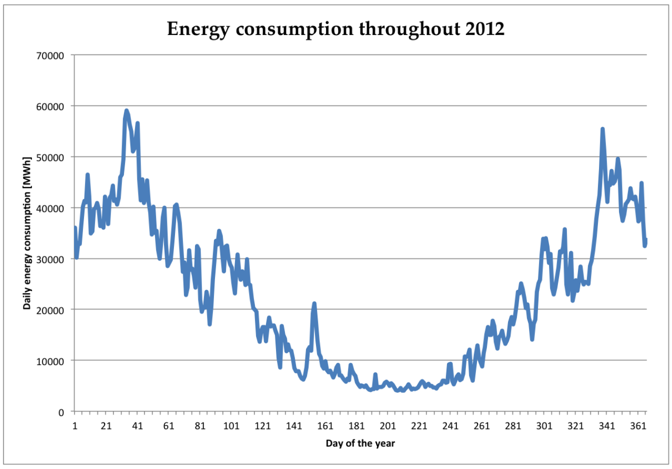


FIGURE 4.12: Daily energy consumption for 2012

A very interesting graph is produced by adding the daily average temperature, but on a negative scale (see figure 4.13). By choosing the scale accordingly, the majority of the observations can be made to coincide, pointing to the fact that the outside temperature has a large influence on the total energy consumption.

This coincidence is less during summer, where the average temperatures increase above about 15 °C. On these days, the consumption does not decrease any longer. The reason for this is that a major part of the buildings does not need heating at these temperatures, but the consumption of hot water continues as usual.

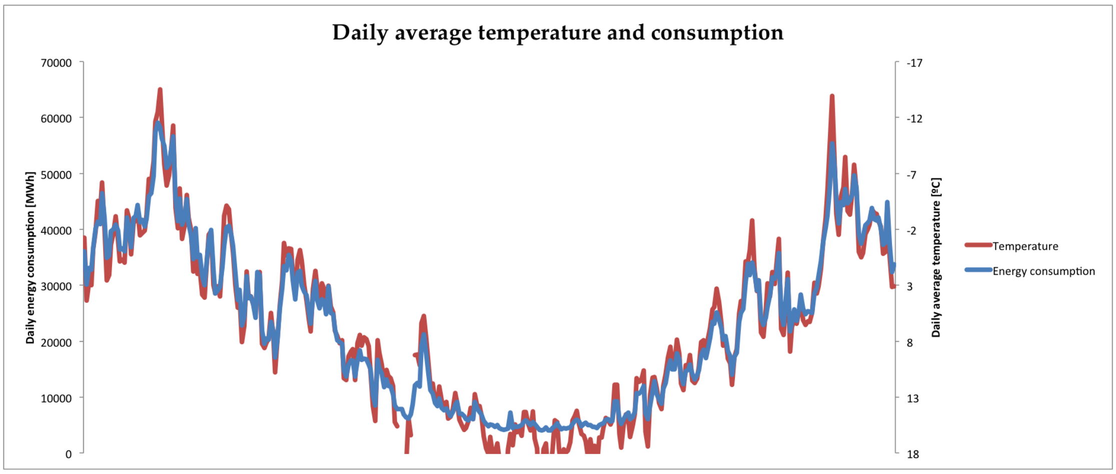


FIGURE 4.13: Daily energy consumption and average temperature

It is however also interesting to see how the consumption varies in the course of one day. Figure 4.14 shows a 3D-plot with the days of the year on one axis, and the hours per day on the other. Cross-sections of the graph along the hour-axis thus show the hourly consumption for one day.

From this figure, it can be seen that most of the days have a peak in consumption during the morning (around 8 AM), while the consumption is on its lowest during nighttime. A remarkable peak during three hours around the 190$^{\text{th}}$ day can be observed. Closer investigation has shown that there is one residential building with a consumption of more than 800 MWh per hour at this time, which coincides with the peak. Since it is the only building with such high consumption around that moment it is assumed that these measurements are flawed.

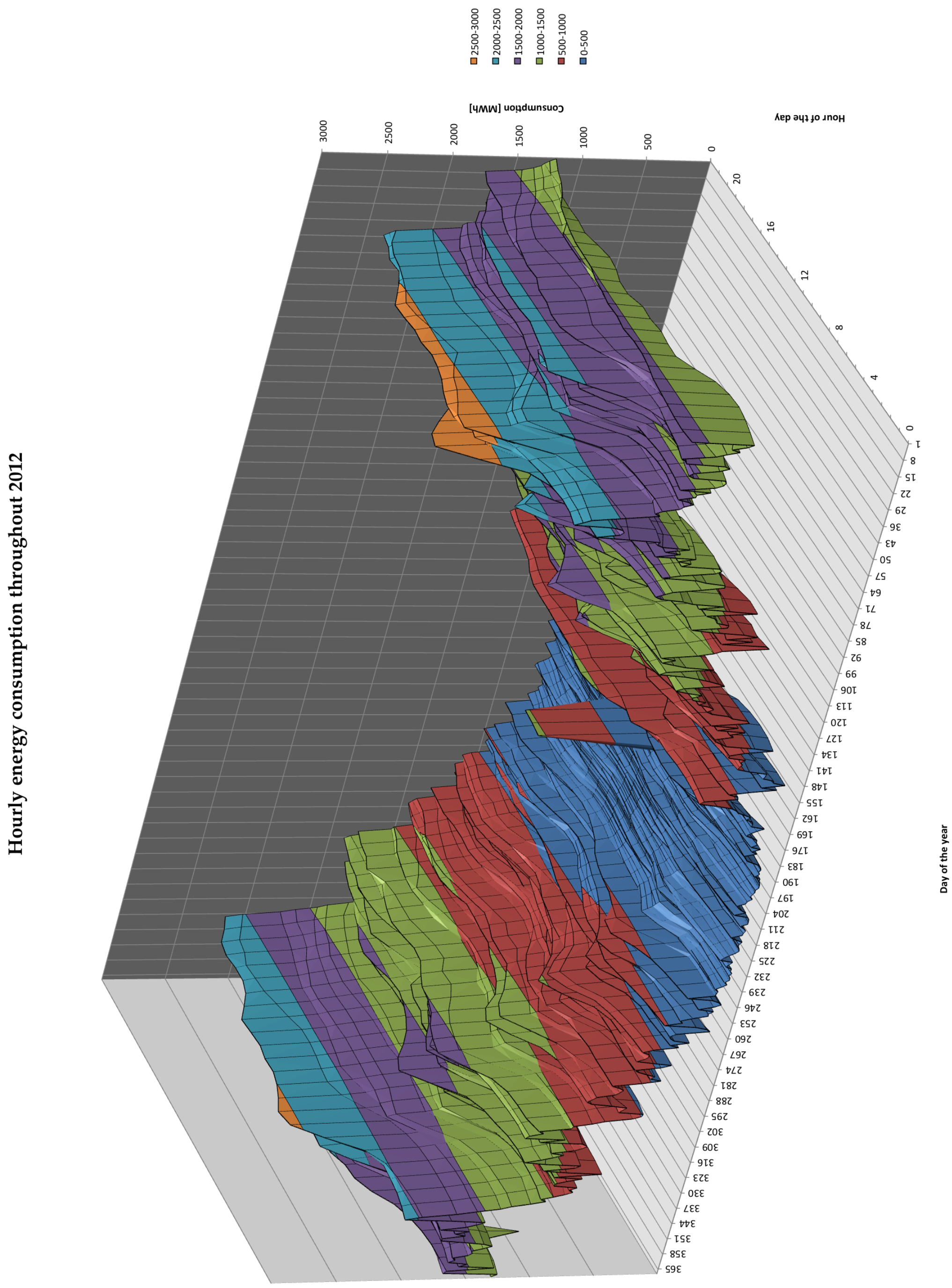


FIGURE 4.14: Hourly consumption for every day in 2012
Note the consumption peaks at 8AM and 6-7PM. The remarkable peak around day 190 at noon can be explained by bad data, since there is one building which consumes the largest part of this peak. Along the day axis, the same evolution as in figure 4.12 can be observed.

4.2.2 Comparison of coldest and warmest day

The observations of the hourly energy consumption are now refined by looking more closely at two individual days. More precisely, the extremes are studied. Since the energy consumption is closely related to outside temperature, this variable is used as a criterion: the coldest and warmest day in the year are considered.

The 4th of February was the day with the lowest average temperature in Stockholm in 2012. Although this was not the day with the highest consumption (3rd of February), their difference in consumption is only minor. Figure 4.15 shows a stacked plot for the hourly energy consumption per class on that day. The largest portion is taken up by residential buildings, followed by commercial buildings. Street heat, other buildings and buildings without classification consume only a marginal amount of energy. The daily peaks at 10 AM and 5-6 PM appear clearly. The hourly average power (or equivalently, aggregate hourly consumption) has a maximum at 2.7 GWh/h.

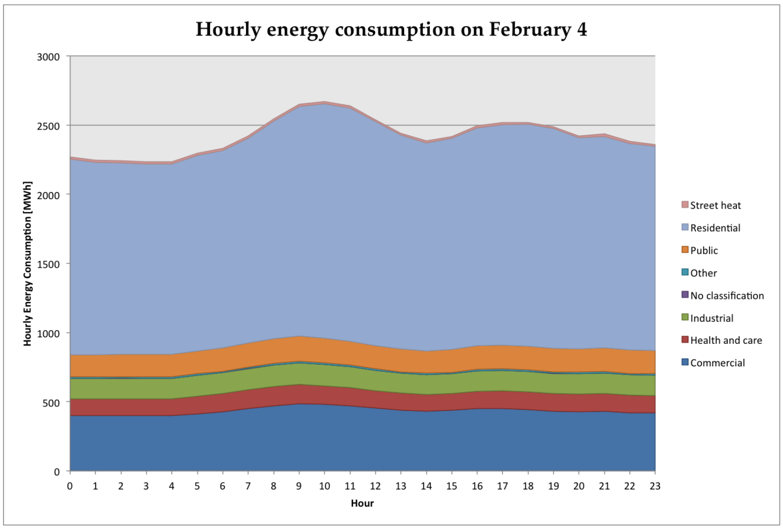


FIGURE 4.15: Energy consumption on the coldest day of 2012

On the warmest day of the year, the 24th of July, the hourly consumption is considerably lower with a peak of only slightly more than 200 MWh at 10 AM. For the rest, the distribution of consumption over the different building classes is similar, with the exception of street heat and no classification, which do not appear in figure 4.16.

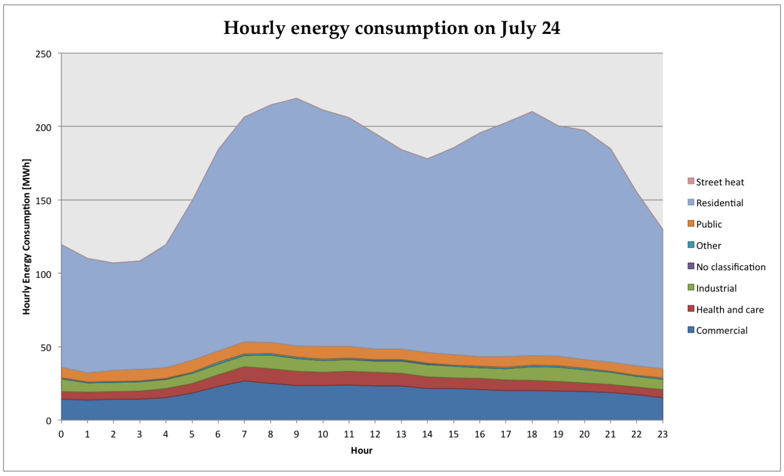
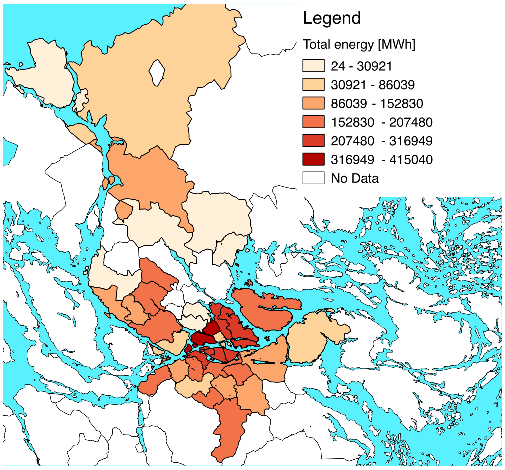


FIGURE 4.16: Energy consumption on the warmest day of 2012

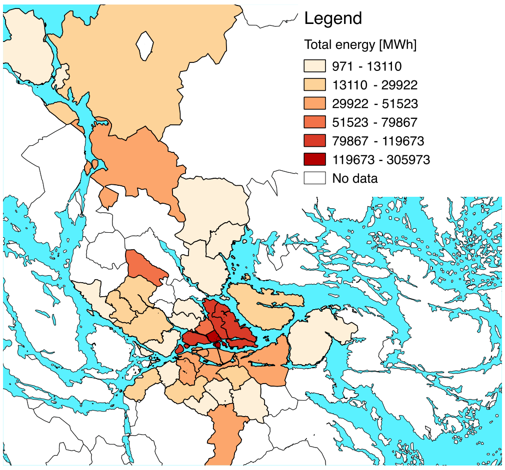
4.3 Energy maps of Stockholm

Figure 4.17 summarises the annual energy consumption data for each 3-digit zip code area. Although the same set of colours is used for all of these maps, one must pay attention to the differing class boundaries. The total annual consumption per class is indicated to provide easier comparison. As expected, residential and commercial buildings have the largest consumption per zipcode area, mostly in and around the centre of Stockholm. For example, the two lowest consumption classes for residential are equal to or greater than the maximal class consumption for most other building classes.

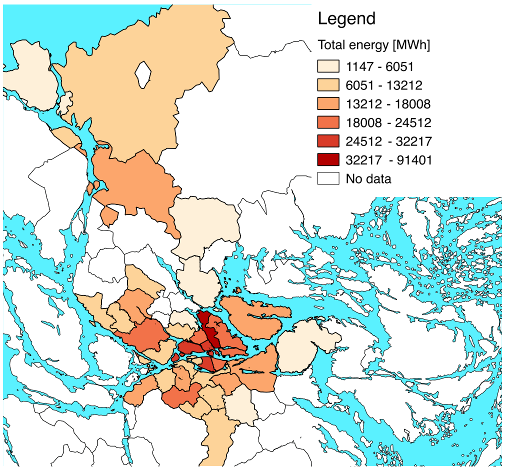
In the first four maps, most of the energy consumption is concentrated in the central areas of Stockholm. The contrast between the centre and the peripheric zip code areas is even greater because of the difference in area between them. It is only for the industrial and other buildings that the consumption in more distant areas is larger than in central areas.



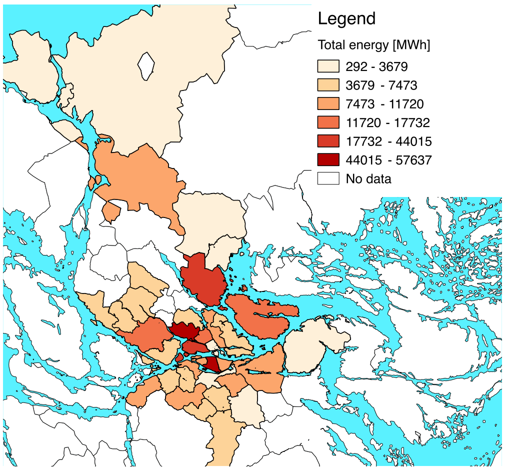
(A) Residential (5,364 GWh)



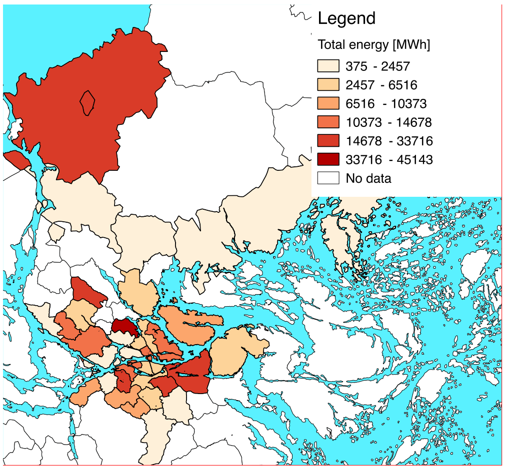
(B) Commercial (1,382 GWh)



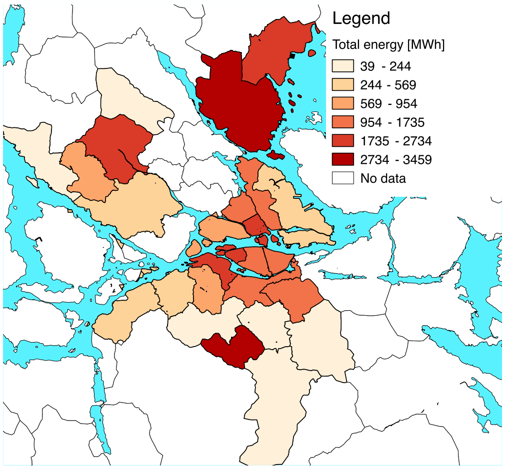
(C) Public (584 GWh)



(D) Health & care (399 GWh)



(E) Industrial (525 GWh)



(F) Other buildings (31 GWh)

FIGURE 4.17: Choropleth maps of total annual energy consumption per building class

4.4 Clustering analysis

This section summarises the most interesting results from the k-means clustering of the various building classes. Where possible, groups of clusters with more or less the same findings are formed. All numeric information about the clusters can be found in appendix C.

4.4.1 Residential buildings

When looking at the clustering results, it is conspicuous that the differences in (average) time of use – both intra-day and intra-week – are very small. Only in some very small clusters (6 and 11), there is a clear deviation of the normal intra-day distribution (peaks at 11-14 and 15-18). Further, clusters that display lower EUIs (below 100 kWh/m^2), there is a tendency towards higher (1, 11 and 19) and lower (17) consumption during weekend days. The remaining clusters appear to have more or less the same distribution in time.

Focusing on the EUI values, there is a gradual increase in EUI across the different clusters. Some values (approximately) occur twice. The difference between these clusters does not appear to be so much in the type of residential buildings, but rather in their building area or total annual consumption.

When we look at the contributing types of housing, the same trend as in the entire stock remains apparent: most of the clusters have a high proportion in detached multi-family buildings, a smaller proportion in houses in a block, and only a fraction of the other types. However, a slight dominance in detached multi-family buildings over buildings that are "attached" seems present for the buildings with a higher consumption (measured in EUI). When the annual consumption cluster averages are sorted ascendingly, this trend seems even clearer in the higher consumption area.
The vintage compositions for all clusters are very different, seemingly irregular. But when the cluster data is sorted according to ascending EUI, the fraction of buildings from 46-75 seems to increase with increasing EUI. On the other hand, the clusters with average EUI below 110 seem to have a higher number of buildings from 1976-2005. These clusters are also the only ones with a noticeable fraction of buildings from after 2006.

4.4.2 Commercial buildings

When the commercial buildings are studied, it can be seen that there are only two clusters with a very small number of observations. Compared to the previous building class, and realising that the amount of commercial buildings is much smaller, we can say that the clusters for the commercial buildings are more evenly divided.

Looking at the time distribution, it can be seen that the clusters with lower EUIs have a somewhat higher consumption during daytime (0, 1, 5, 12, 17). The same can be said for cluster 2, which however shows a higher EUI. The weekly distribution again shows little variation. There are a few clusters which display a lower consumption during the weekend days, but this is hardly noticeable.

Some greater differences in category composition can be noticed for the commercial buildings: although the *Offices & Shops* category is clearly the best represented one, a few clusters have a different composition.
For example, clusters 8 and 11 show a greater proportion of hotels. Coincidentally or not, these clusters have the most evenly distributed consumption per day. Further, cluster 4 and 18 have a remarkably high *warehouse* fraction. However, not much difference with the other clusters can be observed with respect to energy use or temporal distribution.

When the cluster average annual consumptions are sorted from smallest to largest, the fraction of shops and offices seems to increase with the consumption. Comparing the vintage fractions with these EUI values, there seems to be no clear evolution.

4.4.3 Public buildings

In this case, 15 clusters were chosen because of the smaller number of public buildings in the data set. The increase in average EUI per cluster seems to be much larger than in the previous two classes; in addition, the weekend-weekday distribution is more outspoken, in the sense that clusters mainly consisting of schools have an overall lower consumption during weekends. For the intra-day division of the consumption, not much can be said.

Interestingly though, the clusters with the lowest EUI per cluster (up to 142 kWh/m^2) seem to be mainly consisting of schools, while the larger consumers are mostly sports facilities. There are a small number of sports facilities with a normal EUI value (cluster 11). Although the EUI is higher for sports buildings, their annual consumption seems to range from high to low, leading to conclude that there is high potential for efficiency improvements. The small church fraction seems to be more or less equally spread over all clusters.

4.4.4 Health and care buildings

Also for the health and care buildings 15 clusters were derived from the data set. A large number of clusters consists mainly of hospital buildings, just like the composition of the entire class (see 4.8).

Again, only small variations in time distribution seem to be present. However, the clusters that consist of daycare centers for more than 50% have lower consumption during weekends and nights. In addition, the cluster with the highest EUI is found to consist mainly of daycare and recreation buildings. However, there are only 6 small buildings, so there is not much room for efficiency improvements in this case.

On the other hand, clusters that consist of hospital buildings for more than 50% appear to be in the lower range for EUI. Combined with their high building surface and thus annual consumption, this leads to conclude that the health and care buildings perform well overall. Only cluster 2 and 14 (which also has a large service flat fraction) have an EUI of more than 160 kWh/m^2, enabling large efficiency improvements. Service flat clusters (more than 40%) mostly have average EUIs of around 110 kWh/m^2, which is not bad, but according to the new Boverket regulations allows for some improvement.

4.4.5 Industrial buildings

For the industrial buildings, again 20 clusters were constructed. The compositions of these clusters is rather different, and no real patterns can be found. However, a number of clusters have interesting characteristics:

Cluster 1 *Daytime/week day industry.* Contains only three buildings. High EUI of 192 kWh/m^2. 80% of consumption between 6 and 14h. Very little consumption during weekend.

Cluster 3 Seven buildings. Smaller consumption before 10h. High EUI of 233 kWh/m^2

Cluster 7 *Large-scale consumers* Three buildings. EUI of 4131 kWh/m^2, with a maximum value of over 12,000. Consume 170 GWh annually combined, leading to very large savings to be achieved (see remark).

Clusters 4, 9 and 18 *Large industry* About 150 buildings with an EUI of 160 kWh/m^2, with higher annual consumption (larger building area), consuming 194 GWh per year combined.

Cluster 5 *Night industry* Consumes mainly during nighttime, and has a larger than expected consumption during weekends. EUI amounts to ca. 1300 kWh/m^2, and annual consumption for the entire cluster is 13.7 GWh per annum.

Clusters 12 and 14 Their EUIs are 150 and 164 kWh/m^2 respectively. These two clusters combined consume 46 GWh per annum.

Remark: again, the clusters that have high EUI values could also be using the DH system as a process input. In this case, the anticipated savings would be much smaller.

4.4.6 Other buildings

In this class of buildings, a smaller number of clusters of 10 was chosen. There are hardly any surprises in the weekly distribution of consumption, and except for cluster 2 (50-50 warm water only and T-bana stations) where the consumption during the night is lower, the intra-day spread of consumption is more or less even. For this class, a number of groups of clusters becomes clear:

Cluster 1 Consists for 100% of meters with only warm water, and has a low EUI of 37 kWh/m^2, which corresponds with the fact that the consumption of only warm water is lower.

Cluster 0, 2, 4 and 7 Have low EUI values ranging from 64 to 78 kWh/m^2. Mainly T-bana stations and smaller fractions of warm water and heating only meters.

Cluster 8 and 9 Have average EUIs of 211 and 156 kWh/m^2 respectively. Cluster 8 consists mainly of T-bana stations, cluster 9 consists of one meter with heating only and one thermal plant (not charged). They are grouped because both clusters have a total annual consumption of more than 6 GWh.

Cluster 3, 5 and 6 Small number of buildings with high EUIs (resp. 774, 384 and 509 kWh/m^2). Consists of thermal plants and T-bana stations only. Combined annual consumption of 6.6 GWh.

4.4.7 *Key findings*

- The clusters did not uncover well-defined consumer groups with a specific time of use (time of the day/day of the week)
- The clusters are mainly separated based on various combinations of EUI/annual consumption
- The categories that contain the largest share of consumers also appear in most of the clusters. However, there are a number of clusters with clearly discernible category compositions
- In the classes that have a smaller consumption share, the link between consumption behaviour and building category becomes clearer

4.5 Maximum energy savings potential

This section presents the results of the savings potential analysis. The savings per class are summarized according to a number of scenarios, in which EUI limits are imposed. Afterwards, the distribution of the possible savings over the earlier defined clusters and locations in the city are investigated.

4.5.1 Chosen saving scenarios

As stated in the methodology, the EUI limits for the scenarios are derived from the consumption data itself (Touchie et al., 2013) on the one hand, and from Boverket (2013) on the other. The used limits per class are presented in table 4.1. For the low savings the class median has been selected as limiting value for the EUI, except in the case of the category "other", where the median is already lower than the 80 kWh/m^2 that is stated by Boverket. Remark that the consumption limit for medium savings is higher for residential buildings, in accordance with the building regulations.

	Low savings	**Medium savings**	**High savings**
Residential	*137.7*	90	55
Commercial	*100*	80	55
Public	*133.7*	80	55
Health & care	*127*	80	55
Industrial	*122.8*	80	55
Other	80	*63.1*	55

TABLE 4.1: EUI limits for the savings scenarios per building class in kWh/m^2
Class median values *in italics*

4.5.2 Possible energy savings

Following the method in paragraph 3.6.2, the total energy savings per class and per savings scenario can be calculated. The results are first presented numerically in table 4.2. The first column shows the total consumption per class without savings; the columns to the right list the possible energy savings for each scenario. Table 4.3 shows the savings per scenario as a percentage of the total class consumption, with the consumption share for each class in the first column. Remark that the medium savings include the low savings, and the high savings include both low and medium savings.

From the first table, it can be seen that the savings potential for residential buildings is the highest, in accordance with their initial consumption. However, relatively speaking, the possible savings for residential buildings in the low and medium scenarios are only low to average. Since the commercial buildings already display lower average EUI

	Initial	**Savings scenario**		
		Low	*Medium*	*High*
Residential	5,364,073	602,123	1,891,721	3,177,756
Commercial	1,382,347	193,521	319,179	565,220
Public	584,044	96,585	244,666	342,653
Health & Care	399,363	33,116	117,940	198,310
Industrial	524,637	165,318	262,696	335,862
Other	31,363	12,653	14,915	16,140
TOTAL	8,285,827	1,103,315	2,851,118	4,635,942

TABLE 4.2: Savings potential per building class and savings scenario, energy in MWh

	Consumption share	**Savings scenario**		
		Low	*Medium*	*High*
Residential	64.7%	11.2%	35.3%	59.2%
Commercial	16.7%	14.0%	23.1%	40.9%
Public	7.0%	16.5%	41.9%	58.7%
Health & Care	4.8%	8.3%	29.5%	49.7%
Industrial	6.3%	31.5%	50.1%	64.0%
Other	0.4%	40.3%	47.6%	51.5%
TOTAL		13.3%	34.4%	56.0%

TABLE 4.3: Proportion of energy per class and possible savings per class in percentage

values, the percentage savings potential is lower as well. The best results according to this method can be achieved in the industrial and other building class. One must be careful to draw conclusions for these classes, since they are less homogeneous than the other four classes.

Table 4.4 shows the influence of the vintage parameter for residential and commercial buildings. The general observation is that higher savings are possible in older buildings, and that the savings are overall higher for residential buildings, in accordance with the previous tables.
All in all, if the high savings scenario were to be implemented in all studied buildings, a total of 4.6 TWh per annum or 56% could be achieved. 68% of this amount of saved energy originates from residential buildings.

The potential savings can be visualised for easier interpretation. Since the consumption per class is not equally distributed, a marimekko chart is convenient to use here. This chart is drawn in figure 4.18.
On the horizontal axis, the consumption share for each class is represented. It is again clear that residential buildings take up the largest part of district heating consumption.

HC: Health & Care buildings, *Ind:* Industrial buildings, *O:* Other buildings.

	Consumption share	Savings scenario		
		Low	*Medium*	*High*
Residential				
<1925	15.6%	7.3%	34.7%	59.6%
1926-45	19.3%	15.0%	41.4%	63.9%
1946-75	43.2%	13.6%	38.7%	61.7%
1976-	0.02%	9.8%	27.9%	49.5%
1976-05	17.6%	5.5%	22.7%	49.5%
2006-	1.56%	4.3%	13.8%	37.2%
NA	2.69%	9.8%	35.1%	59.5%
Commercial				
<1925	19,3%	20,5%	31,5%	50,0%
1926-45	11,4%	15,9%	25,0%	41,9%
1946-75	26,4%	12,6%	21,3%	39,7%
1976-05	27,2%	8,2%	15,5%	32,6%
2006-	0,6%	3,4%	3,6%	6,5%
NA	15,1%	17,7%	28,6%	46,9%

TABLE 4.4: Proportion of energy per vintage and possible savings per vintage in percentage, both for commercial and residential buildings

On the vertical axis, the percentage of savings per class are represented for each scenario. In this case, to get the total savings for e.g. the medium savings scenario, the upper two areas must be added. In other words, each area for the savings scenario shows how many savings are added with respect to the scenarios with less savings.
An advantage of this type of chart is that the savings per scenario and per class can be compared using the area of each rectangle, which is proportional to the amount of energy that is saved. For example, it is clear that the total savings for the high savings scenario in public buildings is smaller than the savings for the low savings scenario in residential buildings.

4.5.3 Savings per cluster

Figure 4.19 summarises the savings that can be achieved per cluster for every building class. The clusters have been ordered by decreasing energy savings for the respective low savings scenarios. This order was chosen in order to identify which clusters can provide the largest low-cost savings.
In the graphs, the energy that is not saved in either of the scenarios is displayed on the negative y-axis, in order to easily identify the maximum savings on the positive axis.

For the *residential* buildings (4.19a), it is remarkable that the cluster with the largest low savings only has a rather low potential for high savings. At the same time, the energy that cannot be saved in this cluster 9 is remarkably lower (relatively speaking) than in other clusters. Further, it is seen that clusters 10 and 15 have the highest overall

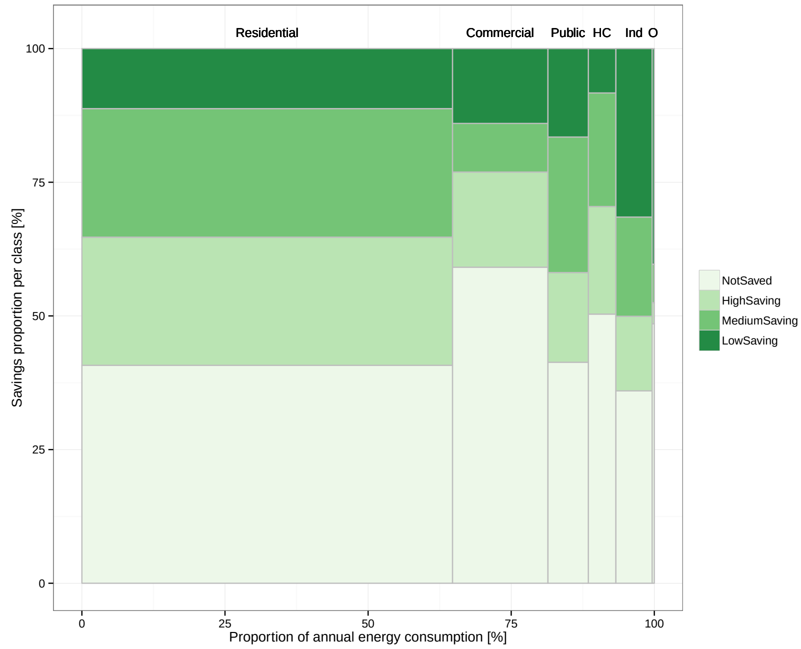
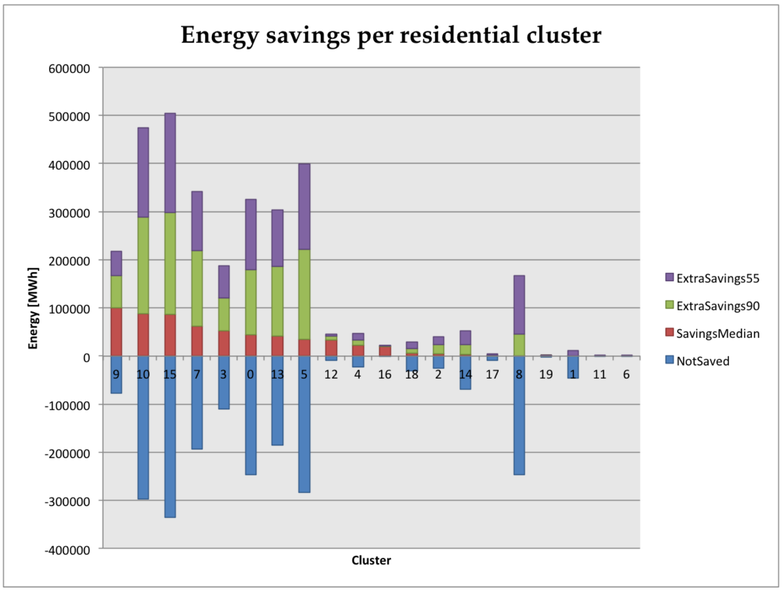


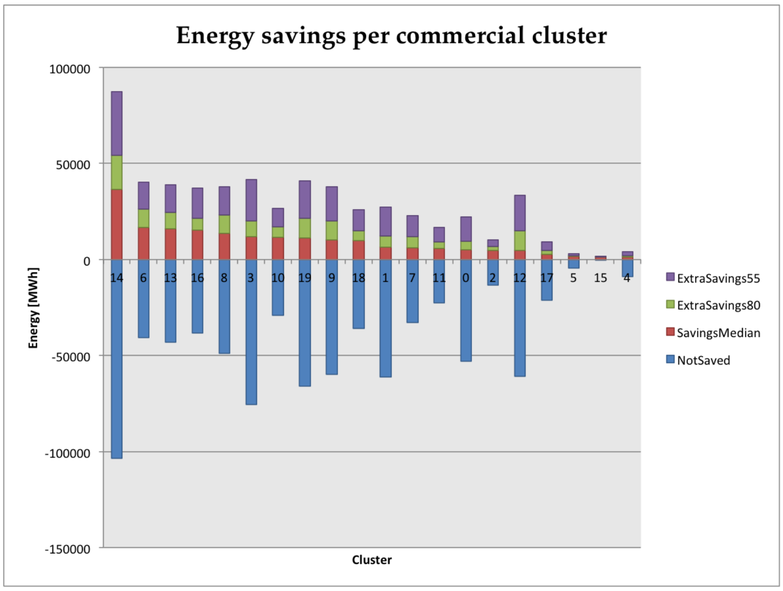
FIGURE 4.18: Marimekko chart of potential savings per class

potential for savings. Since the residential class is the largest consumer in the studied set of buildings, also the savings per cluster are by far the highest.
For *commercial* buildings (4.19b), there is one outstanding peak in cluster 14, of which the potential for low (and thus inexpensive) savings is as high as the total potential for the subsequent clusters.
In the class of *public* buildings (4.19c), the first cluster (9) has both the largest savings potential in absolute as well as in relative numbers (with a small portion of remaining consumption). In spite of the smaller total potential, almost all of cluster 8's energy can be saved with the low savings scenario, making it an interesting cluster for investments. Both previous clusters mainly consist of sports facilities. For the remaining clusters, the easily achievable savings are rather small. The clusters with higher potential seem to encompass mostly school buildings.

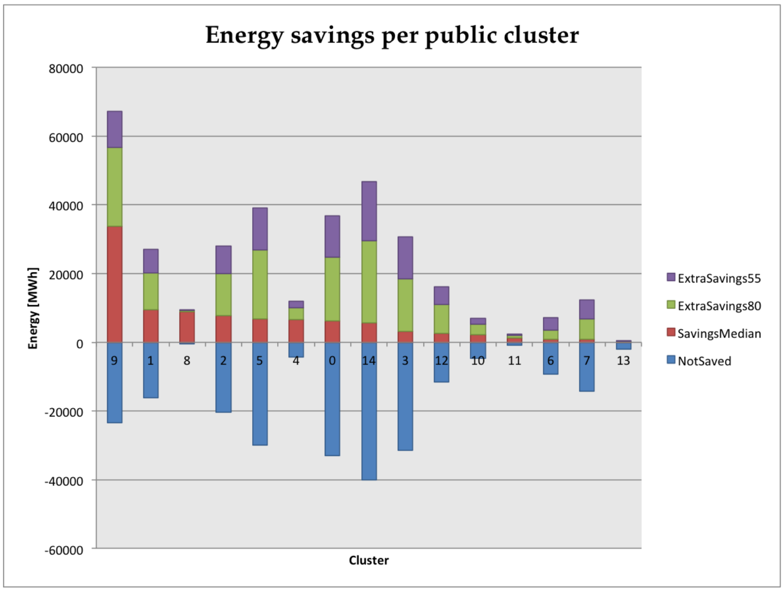
In figure 4.19d it can be seen that *health & care* buildings perform very well already, since the low savings scenario only shows substantial savings for the first three clusters (14,2 and 13). These clusters mainly consist of service flats and hospital buildings. For the remaining clusters, 3 and 10 have the highest savings potential, but they are the highest consumers as well. Further, these savings can only be achieved using the medium and high savings scenarios.
The *industrial* buildings (4.19e) display the same type of behaviour as the commercial buildings, where one cluster (7) has higher low savings than the total potential for the next cluster. This result was already anticipated in the cluster analysis. Although the savings for most of the other clusters seem at the lower end, it must be noticed that



(A)

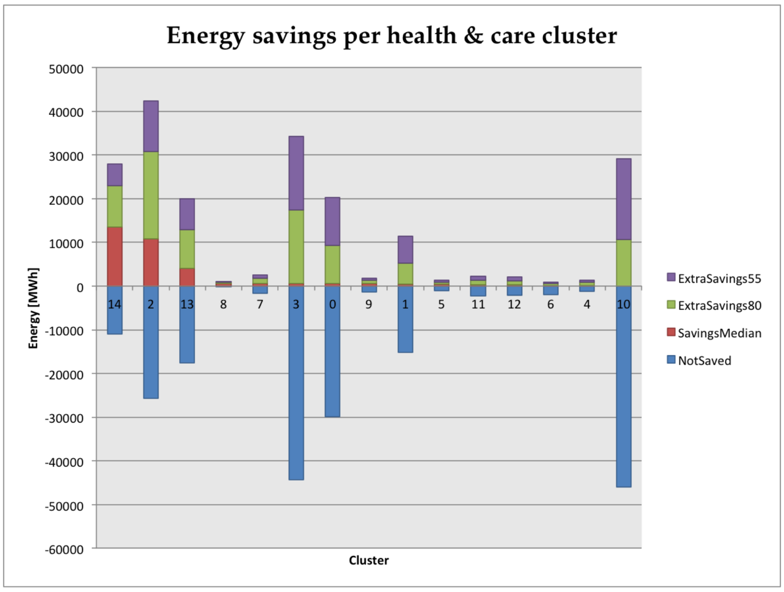


(B)

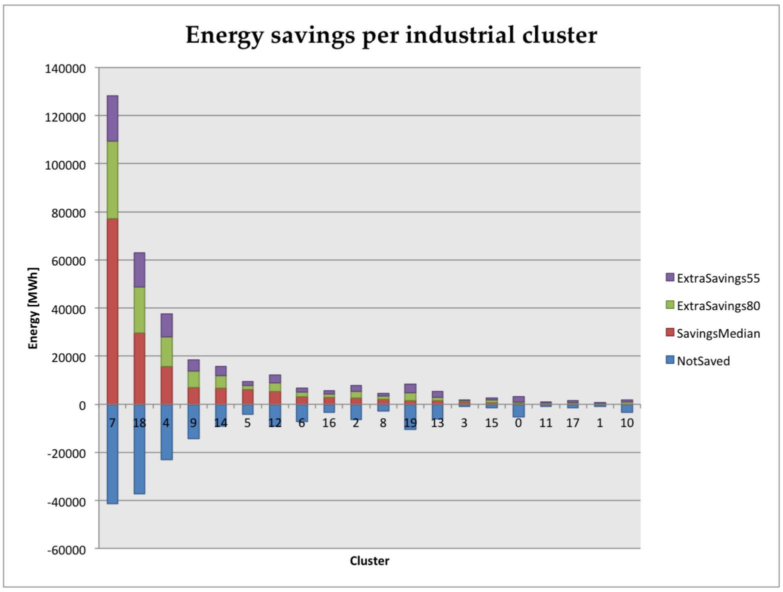


(C)

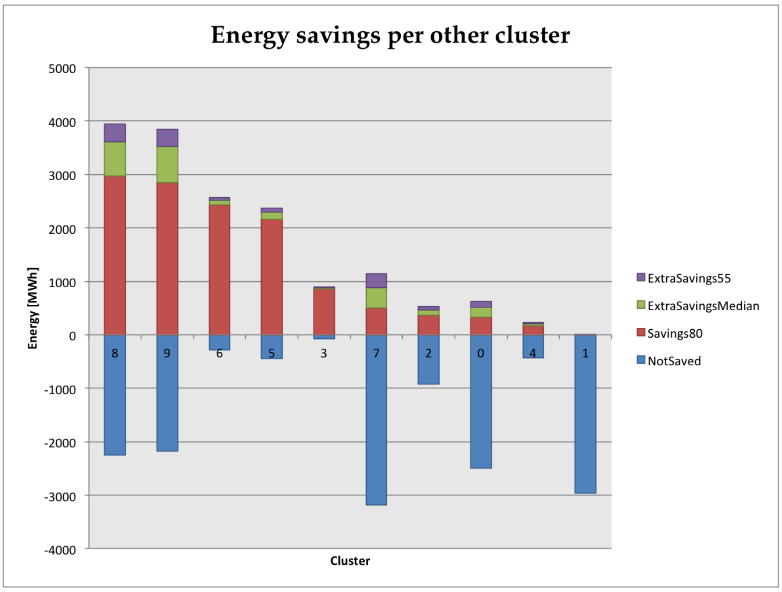
FIGURE 4.19: Cluster prioritisation of energy savings, ordered by low savings



(D)



(E)



(F)

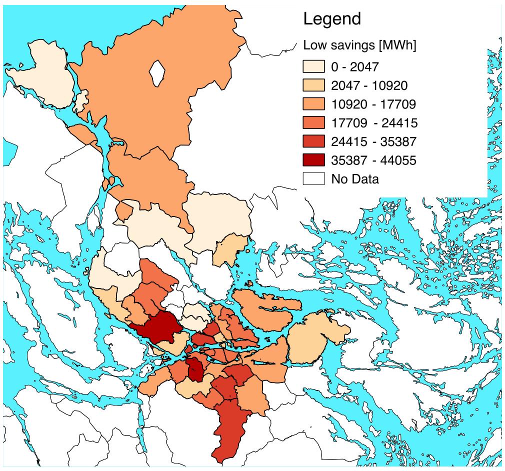
FIGURE 4.19: Cluster prioritisation of energy savings, ordered by low savings *(continued)*

the energy axis shows a much larger scale than most other building classes. Still, the largest savings can be achieved in the first 3 or 4 clusters, certainly when looking at the ratio of saved versus non-saved energy.
Last, the *other* buildings (4.19f) again show a large potential for the low savings scenario. Clusters 8 and 9 have the highest energy potential, seen that even a low savings scenario would save more than in the rest of the clusters all-together. Nevertheless, the relative savings are higher for cluster 3, 5 and 6 (T-bana and thermal plants). For the remaining clusters, only minor relative savings can be realised.

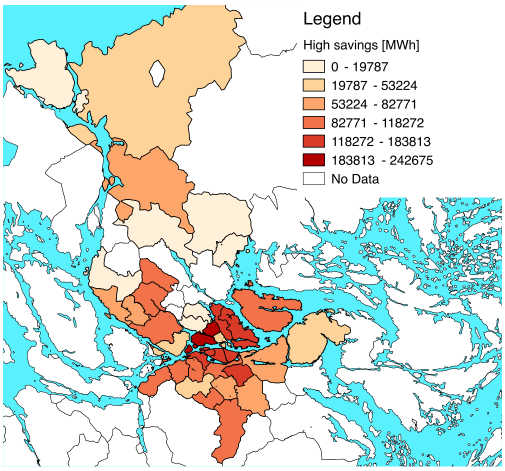
4.5.4 Savings maps

The possible energy savings for two scenarios (low and high savings) are displayed in figures 4.20 to 4.25. Although the same color scales are used in all of the maps, the class boundaries are different for each map. The purpose of the maps is to show that the savings tend to be concentrated in well-defined geographic locations per class.

For residential and commercial buildings, it appears that the higher the savings measures are, the more higher savings seem to occur is a smaller area. This can be seen from the lighter colours towards the edges of the map in the high savings scenario.
This trend seems to be less present in the remaining building classes. For the health & care buildings (figure 4.23), the highest savings appear to be at the location of the city's larger hospitals, such as Karolinska Institutet. Here, the same conjecture as for industrial buildings arises: namely, that heat from the DH network might be used for other purposes than space heating and warm water alone, e.g. disinfection of medical instruments. Further, the class of industrial buildings is the only one to show darker colours in the northernmost parts of Stockholm.

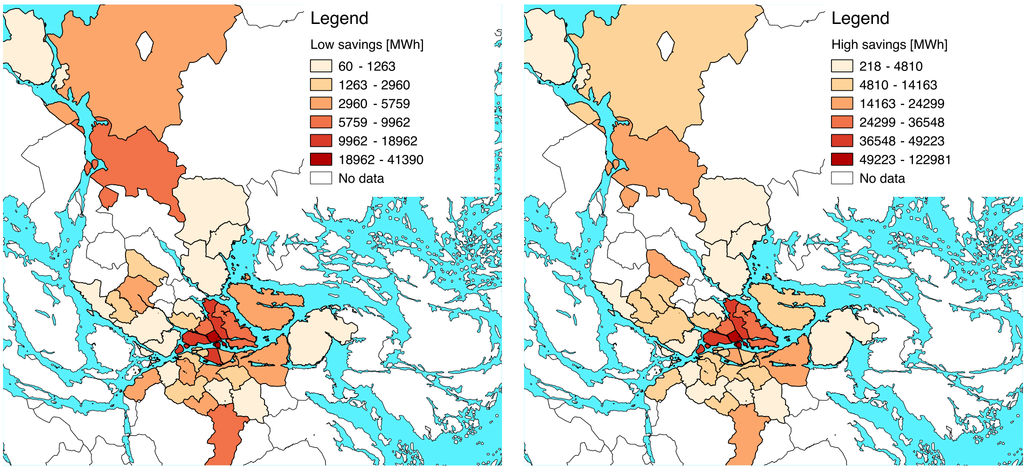


(A) Low savings (*Total:* 602 GWh)



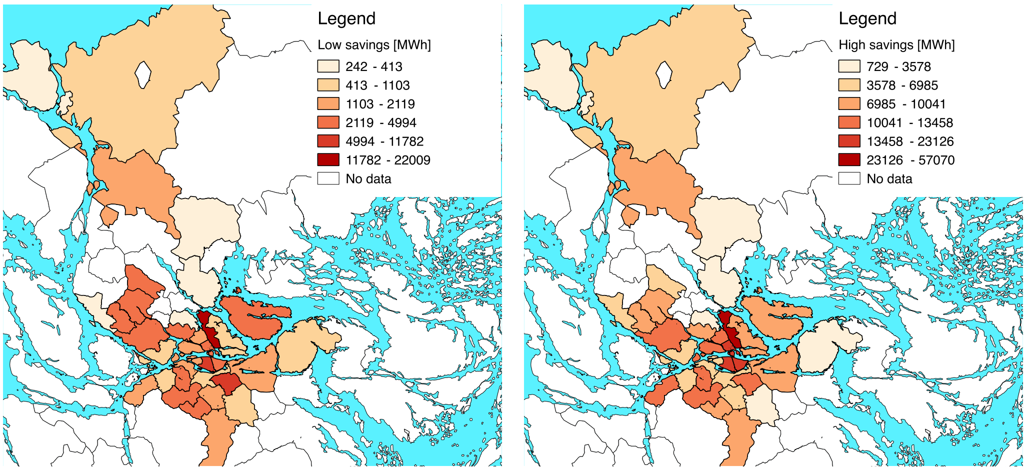
(B) High savings (3,178 GWh)

FIGURE 4.20: Potential energy savings maps for residential buildings



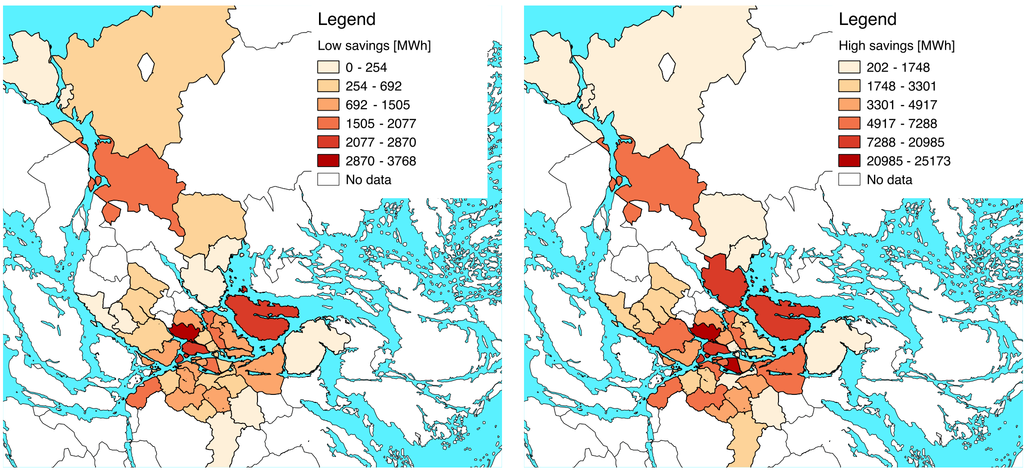
(A) Low savings (194 GWh) (B) High savings (565 GWh)

FIGURE 4.21: Potential energy savings maps for commercial buildings



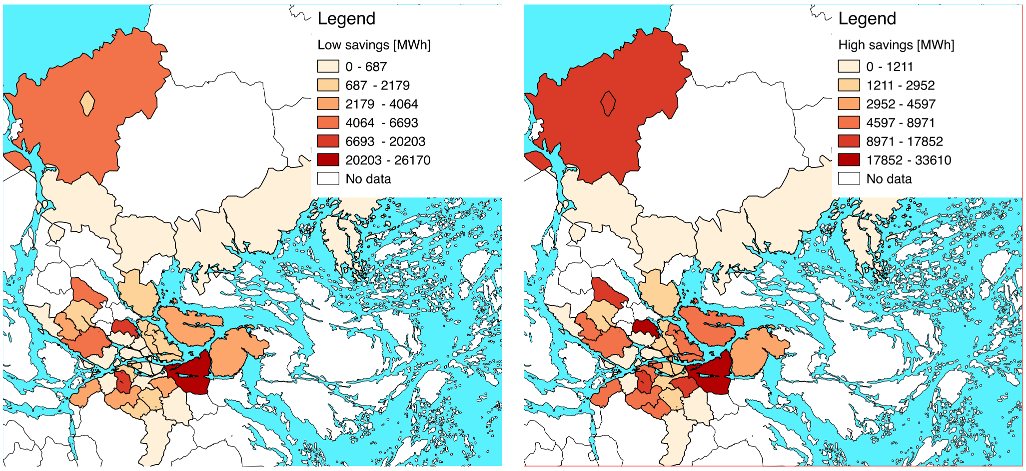
(A) Low savings (97 GWh) (B) High savings (343 GWh)

FIGURE 4.22: Potential energy savings maps for public buildings



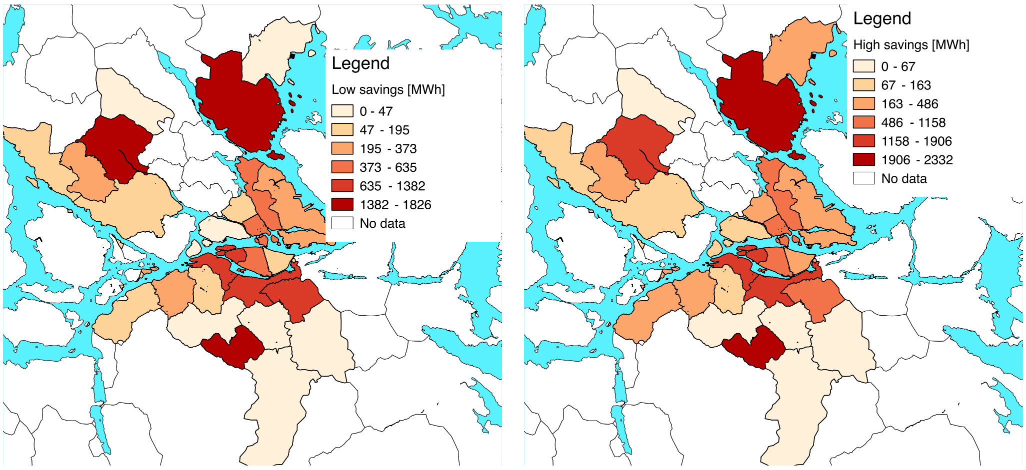
(A) Low savings (33 GWh) (B) High savings (198 GWh)

FIGURE 4.23: Potential energy savings maps for health & care buildings



(A) Low savings (165 GWh) (B) High savings (336 GWh)

FIGURE 4.24: Potential energy savings maps for industrial buildings



(A) Low savings (13 GWh) (B) High savings (16 GWh)

FIGURE 4.25: Potential energy savings maps for other buildings

4.5.5 *Key findings*

- The maximum energy savings potential is estimated at 59% of the annual consumption, or 4.6 TWh. This corresponds to the scenario in which all buildings have an EUI of 55 kWh/m^2 or less.
- The largest share of these savings (about 70% of the 4.6 TWh) are achieved in residential buildings.
- If the savings are compared per class, the residential, public and industrial classes perform best. Each of them can save up to 60% of the initial consumption.

Note: as indicated earlier, it is not sure whether the industrial consumption can be attributed to heating and hot water only. These savings numbers must be interpreted carefully.

- Where available, the building age factor indicates that higher average savings are achieved in older buildings.
- Both savings maps and building clusters serve as a tool to identify buildings that enable the highest savings and thus to target cost-efficient investments in energy efficiency.

4.6 Retrofitting plan

This section presents the results of the retrofitting analysis in order to achieve the City of Stockholm's goal of 5% building energy savings from 2012 to 2015. A investment horizon of maximum 15 years was chosen. The costs of renovation were chosen at 50 kr/m^2 (small savings), 700 kr/m^2 (HVR system) and 2000 kr/m^2 (building shell renovation). As pointed out earlier, industrial buildings are not included in the analysis because of their high consumption variation.

4.6.1 Goal savings

In this first case, a retrofitting plan was constructed that achieved annual energy savings of 5% on the input set of buildings (without industrial). The plan consists of 834 buildings (out of 14 119), that all have a positive ROI. Out of these buildings, 728 will undergo a building shell renovation, while the remaining 103 need a new heating control system (small savings). No buildings are retrofitted with a HRV. The total investment needed is 3.23 billion kr ($3.2304 \cdot 10^9$), whereas the saved energy would have costed 5.25 billion kr in the planning horizon of 15 years. This corresponds with annual energy savings of 389 GWh.

Figure 4.26 shows the distribution of the energy use intensities in the buildings from the retrofitting plan. Most of the buildings to be retrofitted appear to have EUIs between 200 and 300 kWh/m^2, but there are a number of exceptions with much higher, but also lower EUI values. In figure 4.27, the respective energy savings per building category (grouped according to the five building classes) are shown.

In terms of maximum power, the 5% saving scenario would reduce the peak demand on the coldest day of the year with 114 MWh/h (was 2 500 MWh/h). The reduction of the demand profile for this day for the different calculation cases is visualised in figure 4.28. The relevance of this result lies in the reduction of the peak heating load; GHG emissions can be decreased more than proportionally if low efficiency peak plants don't have to be started up. This will also decrease the total running cost for the DH system, since these peak plants are powered with more expensive fuel.

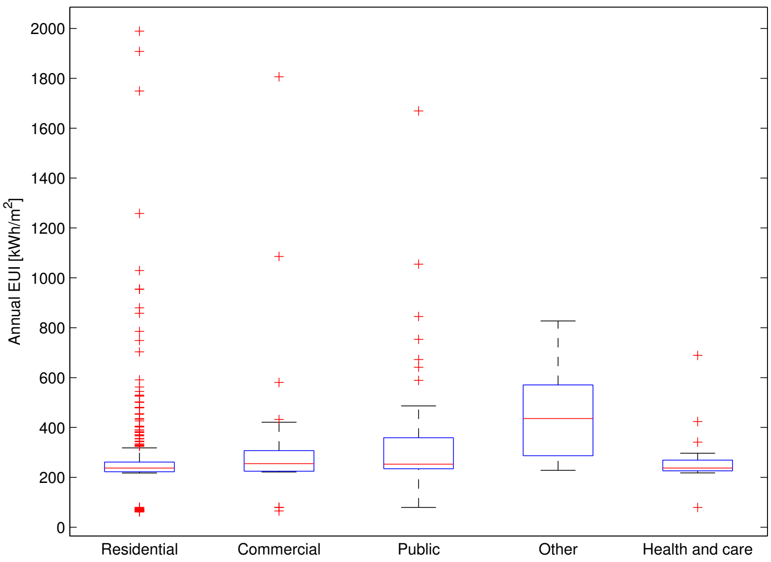


FIGURE 4.26: Boxplot of annual EUI for the 834 buildings in the 5% energy savings retrofitting plan

4.6.2 Maximum profitable savings

With the assumptions about the calculation variables as stated above, even larger savings than the 5% goal can be achieved while still maintaining a positive ROI for every individual building. According to the savings simulation, 1277 buildings could be retrofitted, yielding an annual consumption saving of 472 GWh, or 6.1% compared to the total consumption for the studied set of buildings. As figure 4.28 points out, this is only a minor difference from the 5% savings goal in terms of peak power savings. The total retrofitting cost is approximately 4.3 billion kr

Remark that this retrofitting plan has an overall higher monetary gain from energy savings than the combined investments. This is because only the last building has an exact break-even situation, whereas all other investments have a pay-back time shorter than 15 years. Most of the buildings in this plan receive a complete building shell renovation (934), while all the rest is retrofitted with new thermostats (342).

The distribution of the savings over the buildings is more or less the same as in the previous case, except for the single family houses, which double their share at the expense of savings for sports facilities. It is also interesting to have a look at the energy savings in the different clusters. Figure 4.29 shows the amount of energy saved per cluster and per building class. In case of the health & care buildings, the prognosis that the highest savings would be possible in cluster 2 and 14 seems to be correct.

It is interesting to see that the highest economical savings do not always coincide with the highest possible savings. For reference, the total savings per class are indicated in figure 4.30.

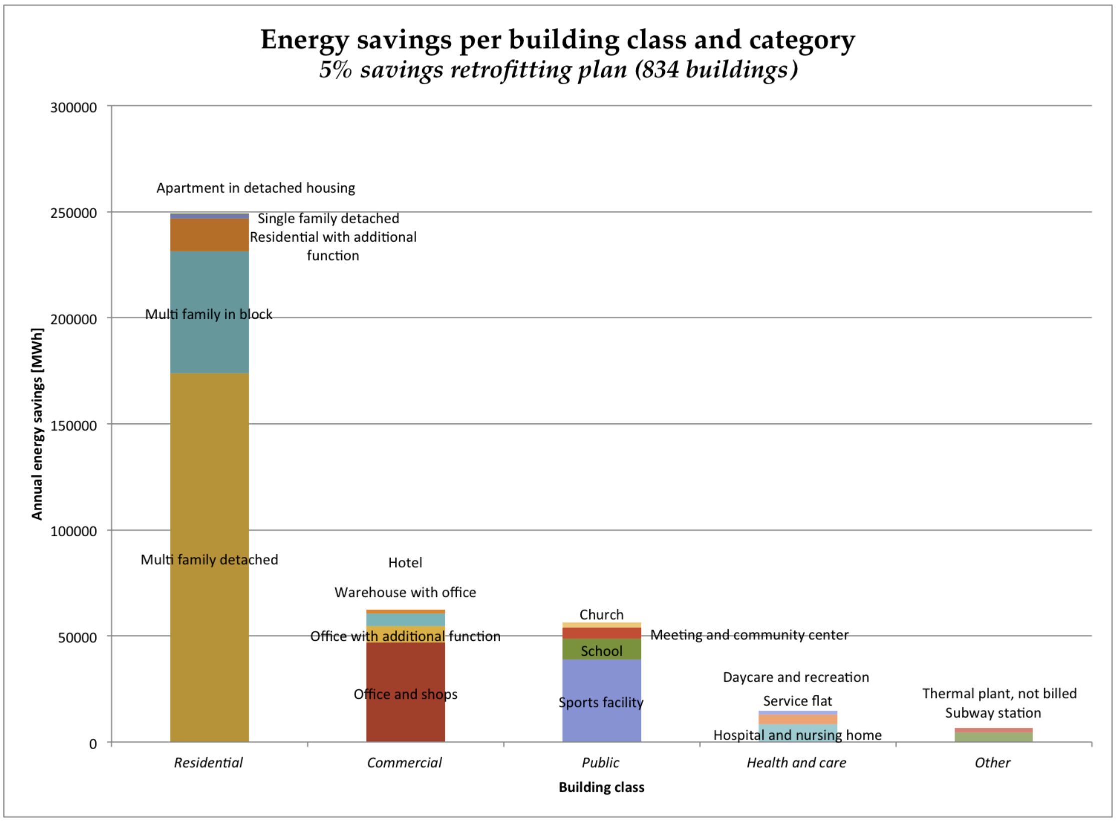


FIGURE 4.27: Energy savings per building category in the 5% energy savings retrofitting plan

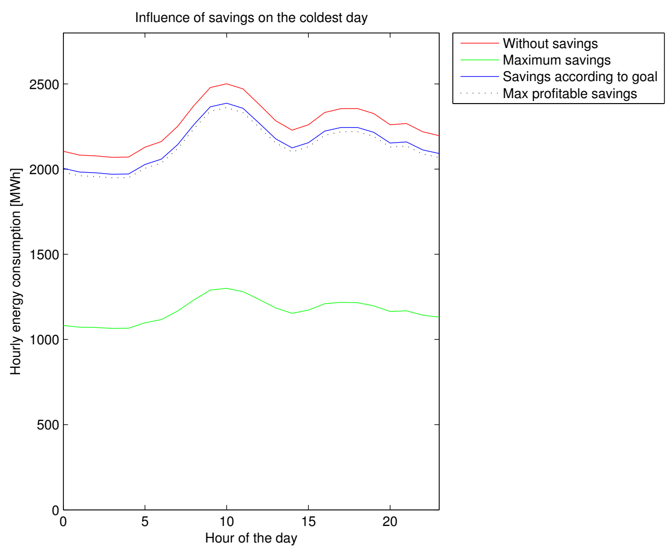


FIGURE 4.28: Heating demand reduction for the coldest day of the year

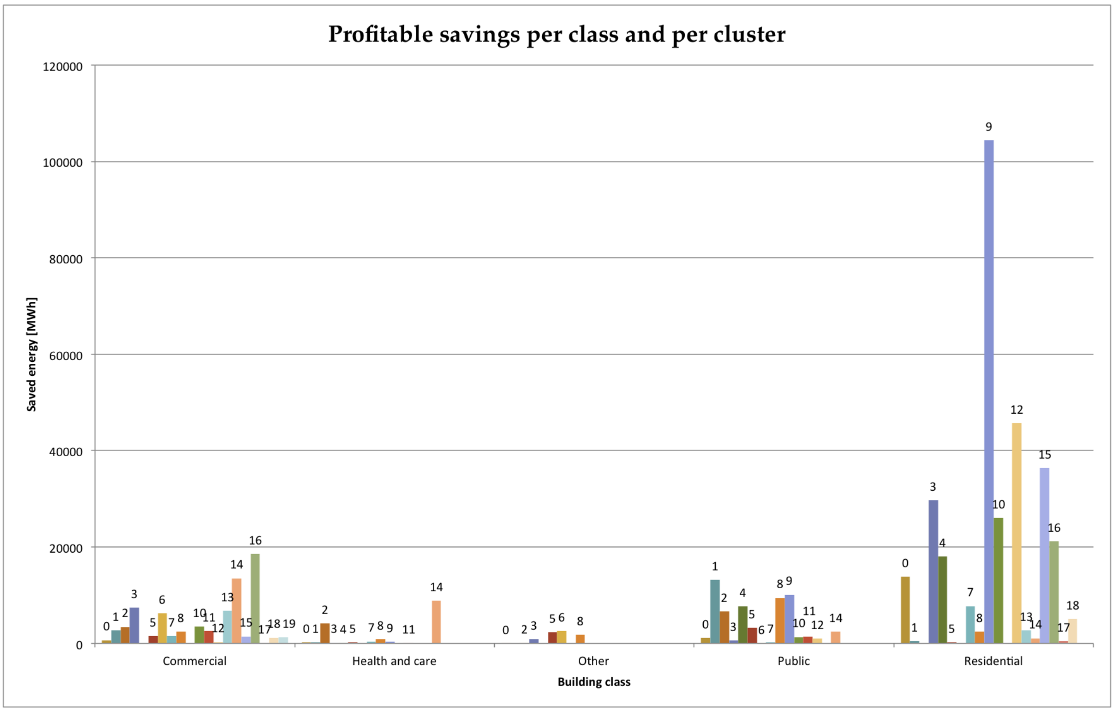


FIGURE 4.29: Maximum profitable savings per building class and cluster

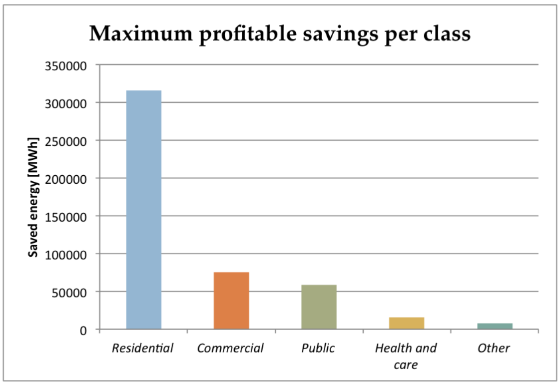


FIGURE 4.30: Maximum profitable savings per class

To finish the results about the maximum profitable savings, figure 4.31 shows a map with the geographical spread of the profitable savings. The largest savings are concentrated in a small number of areas around the city center (Norrmalm & Gamla Stan, Västberga, Södermalm, Bromma and Johanneshov), and one area in the north (Upplands Väsby).

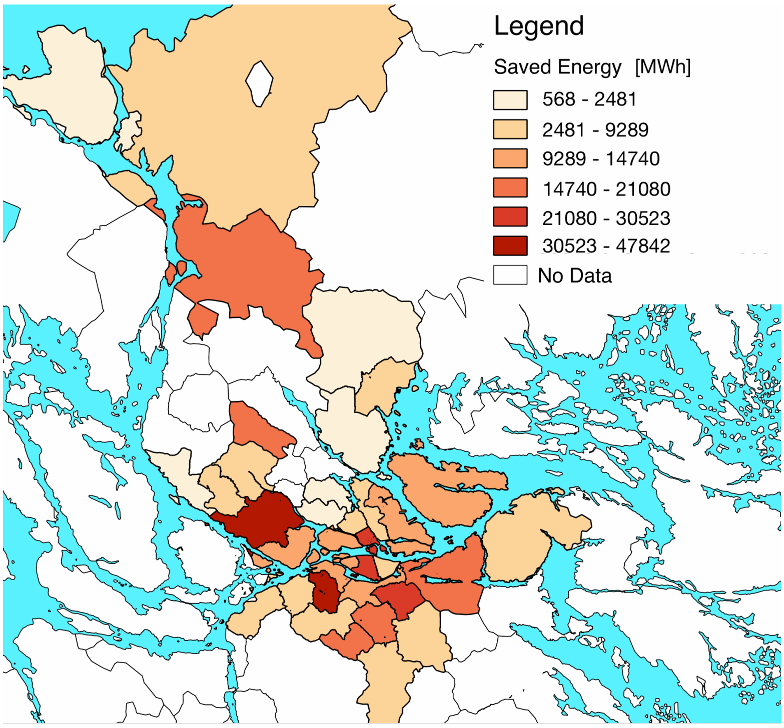


FIGURE 4.31: Map of maximum profitable savings (472 GWh)

4.6.3 Maximum possible savings

The last case that is studied is that in which all buildings that have an EUI greater than 60 kWh/m^2 are retrofitted according to the different types of retrofit measures that are mentioned in the methodology section, irrespective of the economic situation of that investment. In this case, there is an annual savings potential of 3 831 GWh or 49.35% of the total annual consumption (again without industrial buildings). The combined investment that would be needed amounts to 76.9 billion kr, while the combined return because of the saved energy for 15 years would only sum up to 51.7 billion kr.

In this non-economic scenario, 7887 buildings would need a shell renovation, 3396 would be retrofitted with a HRV system, and 576 would receive new thermostats. There are 984 buildings that are below the EUI limit for the retrofitting scenarios. Based on building floor area, 56% of these buildings without savings are commercial, 24% residential, 15% public and the remaining 5% are equally shared by health & care and other buildings.

Again, on figure 4.28 the possible peak demand reduction can be seen; the maximum reduction amounts to 1.2 GWh/h at 11 AM; this reduction is close to 50% of the peak demand.

4.6.4 Sensitivity of the results

In order to assess the influence of the chosen parameters, a limited sensitivity analysis is performed. The first parameter to be investigated is the cost of a climate shell renovation, as the largest investment cost and apparently also the most needed type of retrofitting.

The price of retrofitting the climate shell is varied between 1500 and 3500 kr/m^2. Since the ROI is inversely proportional to the cost of the investment and thus to the unit cost of the shell retrofitting, the maximum profitable savings (i.e. with a positive ROI) as function of the retrofitting price have a hyperbolic shape (see figure 4.32). A similar pattern can be observed in the peak power demand reduction (figure 4.33).

This makes investment in a large retrofit very risky when the price can be assumed to be in the lower range (1500-2500 kr/m^2). If the unit price is higher, the total savings are significantly less (remember that around 400 GWh/y is needed in order to achieve the 5% goal, and that a large part of these savings can be realised by climate shell improvements), but the uncertainty about the result is much smaller.

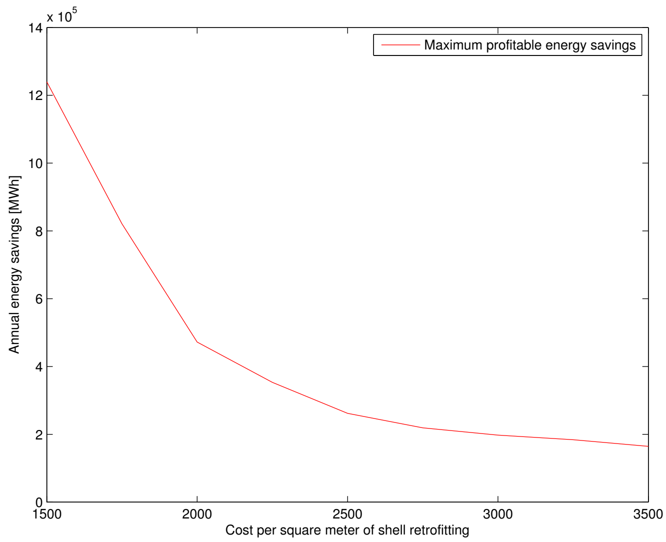


FIGURE 4.32: Dependence of maximum profitable savings on the cost of climate shell renovation

Although the sensitivity to the retrofitting cost is high, the effect is slightly mitigated for the 5% savings goal retrofitting plan by the fact that the less interesting investments move down on the priority list. This effect is well illustrated in figure 4.34. The larger the costs of the high savings retrofit, the less they are implemented and the more they are replaced by (less effective, but cheaper) other retrofitted buildings. Needless to say, these more expensive plans also include buildings with a negative ROI, and the cost still nearly doubles when the case for 1500 and 3000 kr/m^2 are compared.

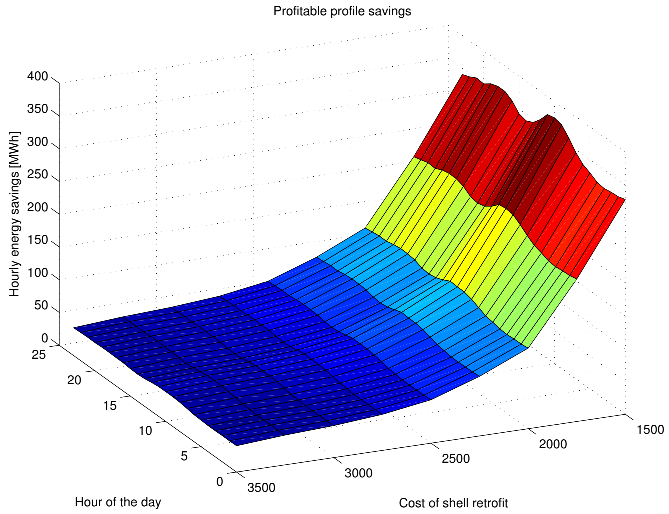


FIGURE 4.33: Power demand reduction for the coldest day of the year as a function of the retrofitting price

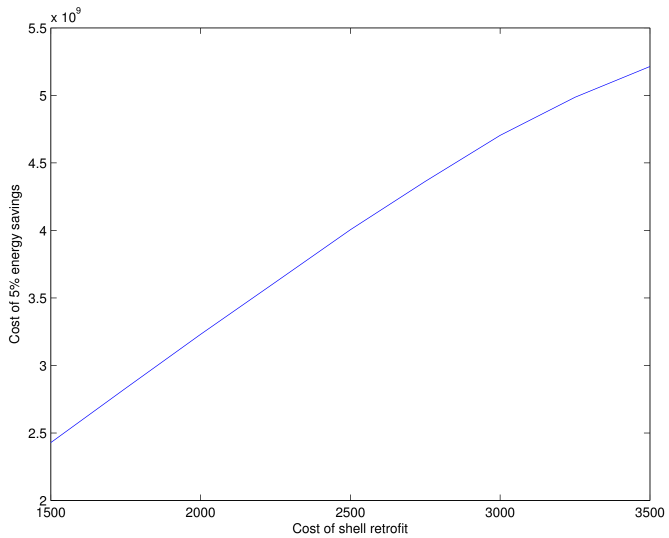


FIGURE 4.34: Total cost (kr) for the 5% savings retrofitting plan

The second parameter that is investigated for sensitivity is the investment horizon. As mentioned before, this parameter is rather uncertain. A too long horizon increases the risk and makes the investment less interesting. A too short horizon makes it impossible to achieve the goals that were set. The base case that was used until now assumed a value of 15 years, which is already quite long, and it would not be very interesting to surpass this value, since the 5% goal can be achieved in this way. Therefore, the horizon was varied between 3 and 15 years with intervals of 2 years.

Figure 4.35 shows the possible savings with a positive ROI as a function of the investment horizon. The possible savings are halved when the horizon is shortened from 15 to 10 year, and the decrease is even larger if the horizon is assumed to be 5 years. More or less the same pattern of decreasing savings can be observed for the peak demand reduction in figure 4.36.

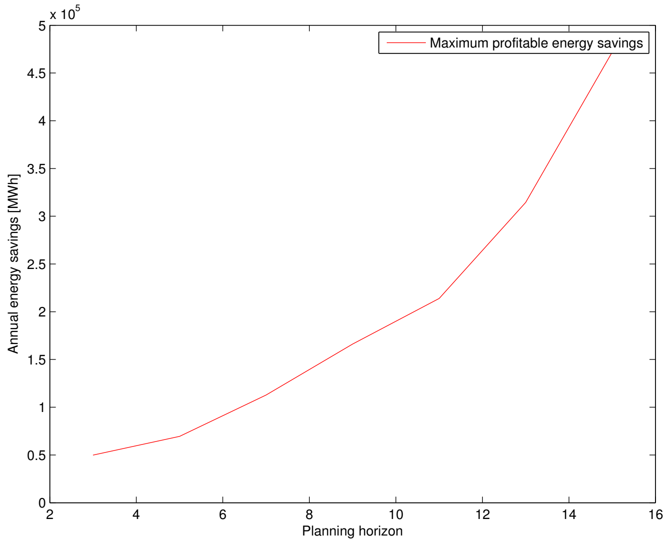


FIGURE 4.35: Maximum profitable savings as a function of investment horizon

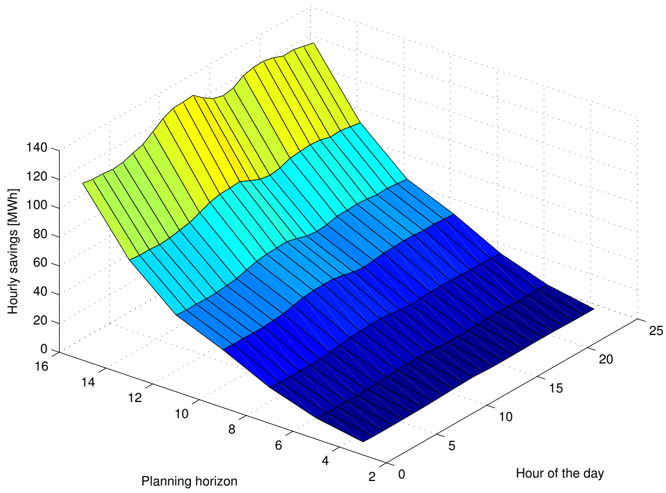


FIGURE 4.36: Power reductions on the coldest day as a function of investment horizon

4.6.5 *Key findings*

- This section has proposed three savings measures that can be implemented:
  - Replacement of thermostats
  - Installing a HVR
  - Complete climate shell retrofit
- Further, the economic aspects of retrofitting have been investigated (installation cost vs. running cost savings)

- There is a total potential of 6.1% (472 GWh per year, 1277 buildings, cost 4.3 billion kr) of energy savings by retrofitting buildings that have a pay-back time of less than 15 years. This means that the Stockholm goal of 5% energy savings in buildings from 2012 to 2015 can be achieved with reasonable investments.
- The total energy savings potential using the proposed measures, irrespective of costs and benefits, is 49% or 3 831 GWh per year.
- Again, maps and clusters can be used to identify groups (both based on location or characteristics) that are particularly interesting from an investment perspective.

4.7 Weather influence analysis

This section summarizes the results of the weather influence analysis on residential buildings, as described in section 3.8. First, the results of the regression on 3-digit zip code level without further classification are presented. Thereafter, classifications based on annual EUI and building type are presented.

4.7.1 Spatial regression

In this analysis, a linear regression of the EUI* with respect to the wind speed and temperature is conducted for each 3-digit zip code separately.

Table 4.5, 4.6 and 4.7 show the regression results for the intercept, temperature and wind speed coefficient respectively. It can be seen that the t-statistic is generally very high in absolute value and in most of the cases higher than the required 1.96 (at significance level 5%) for the coefficients to be statistically different from 0. This is otherwise shown by the P-value, which indicates the probability that a higher value of this t-statistic is encountered if the null hypothesis (i.e., that the coefficient is actually equal to zero) is true. A P-value of 0 (in most of the cases) indicates very strong statistical evidence to reject the null hypothesis. Only when the value is higher than 0.05, the statistical evidence for rejecting the null hypothesis is insufficiently strong.

From the first two tables, it can be concluded that the coefficients for all studied zip code areas are significant to at least a level of 5%. For the wind speed coefficients, a number of coefficients do not have a P-value of 0. In the case of zip code 187 it can effectively be concluded that the null hypothesis (that the coefficient equals 0) cannot be rejected. A second interesting case is zip code 169, which displays – like 187 – a highly unlikely negative linear relation between wind speed and energy consumption. Here, the P-value has risen to 0.018, for which the null hypothesis can still be rejected at

Coeff.	Std. Err.	t-value	P-value	Zip3	R Squared
3.93E-03	2.75E-05	142.68	0	111	0.30
3.94E-03	5.30E-06	742.75	0	112	0.73
3.96E-03	2.88E-06	1372.76	0	113	0.89
3.99E-03	8.70E-06	458.84	0	114	0.53
3.98E-03	8.12E-06	489.76	0	115	0.66
3.90E-03	3.08E-06	1266.67	0	116	0.92
3.95E-03	5.30E-06	744.23	0	117	0.86
3.92E-03	4.34E-06	903.80	0	118	0.87
4.11E-03	8.82E-06	466.31	0	120	0.74
3.97E-03	6.85E-06	579.41	0	121	0.71
4.16E-03	5.43E-06	766.81	0	122	0.81
3.89E-03	1.20E-05	324.52	0	123	0.66
3.99E-03	3.55E-05	112.33	0	124	0.16
4.05E-03	9.24E-06	438.34	0	125	0.75
3.87E-03	1.00E-05	385.93	0	126	0.64
3.97E-03	2.32E-05	171.19	0	127	0.29
4.15E-03	6.92E-06	599.44	0	128	0.82
3.93E-03	2.47E-05	159.30	0	129	0.23
4.00E-03	2.21E-05	181.03	0	131	0.48
4.14E-03	8.73E-05	47.37	0	132	0.13
4.23E-03	9.40E-06	449.63	0	162	0.76
3.91E-03	8.26E-06	472.90	0	163	0.86
3.85E-03	9.94E-06	387.67	0	164	0.83
4.14E-03	1.02E-05	405.69	0	165	0.62
4.13E-03	7.36E-06	561.24	0	167	0.81
4.22E-03	6.04E-06	698.37	0	168	0.78
5.59E-03	1.67E-04	33.60	0	169	0.87
3.95E-03	1.83E-05	215.66	0	171	0.97
3.94E-03	2.42E-05	162.92	0	175	0.98
3.99E-03	6.68E-06	596.59	0	181	0.72
4.12E-03	5.16E-05	79.89	0	183	0.57
4.20E-03	4.06E-05	103.49	0	187	0.88
4.05E-03	7.28E-05	55.64	0	192	0.93
4.14E-03	3.21E-05	129.00	0	193	0.63
4.12E-03	1.51E-05	271.93	0	194	0.60
3.96E-03	7.71E-06	514.43	0	195	0.91

TABLE 4.5: Intercept results by 3-digit zip code for spatial regression

a significance level of 5%, but not at 1%. The other coefficients with a non-zero P-value don't seem to have strange values.

The intercept is not of great interest for this analysis, since it does not tell anything about the dependence of the EUI* on wind and temperature. This dependence is indicated by the temperature and wind speed coefficients.
In terms of temperature dependence, the energy consumption seems to be influenced in more or less the same way in all zip code areas. Indeed, in most of the studied areas, the coefficient is roughly between $-2.4 \cdot 10^{-4}$ and $-2 \cdot 10^{-4}$. Only in the area with zip code 169 (which also had a counter-intuitively negative wind speed regression coefficient) does not comply with this trend and equals $-3.7 \cdot 10^{-4}$. In order to visualize

Coeff.	Std. Err.	t-value	P-value	Zip3	R Squared
-2.21E-04	1.24E-06	-178.74	0	111	0.30
-2.15E-04	2.38E-07	-901.85	0	112	0.73
-2.21E-04	1.30E-07	-1702.24	0	113	0.89
-2.29E-04	3.91E-07	-584.45	0	114	0.53
-2.19E-04	3.65E-07	-598.95	0	115	0.66
-2.12E-04	1.38E-07	-1529.76	0	116	0.92
-2.15E-04	2.38E-07	-902.83	0	117	0.86
-2.16E-04	1.95E-07	-1104.07	0	118	0.87
-2.25E-04	3.97E-07	-568.64	0	120	0.74
-2.16E-04	3.08E-07	-700.51	0	121	0.71
-2.37E-04	2.44E-07	-969.17	0	122	0.81
-2.02E-04	5.39E-07	-374.64	0	123	0.66
-2.14E-04	1.60E-06	-134.08	0	124	0.16
-2.25E-04	4.15E-07	-541.67	0	125	0.75
-2.06E-04	4.51E-07	-455.27	0	126	0.64
-2.10E-04	1.04E-06	-201.49	0	127	0.29
-2.30E-04	3.11E-07	-738.81	0	128	0.82
-2.09E-04	1.11E-06	-188.43	0	129	0.23
-2.15E-04	9.93E-07	-216.95	0	131	0.48
-2.44E-04	3.93E-06	-62.12	0	132	0.13
-2.39E-04	4.23E-07	-566.14	0	162	0.76
-1.98E-04	3.72E-07	-532.54	0	163	0.86
-1.98E-04	4.47E-07	-441.95	0	164	0.83
-2.33E-04	4.59E-07	-508.34	0	165	0.62
-2.38E-04	3.31E-07	-718.05	0	167	0.81
-2.36E-04	2.71E-07	-867.57	0	168	0.78
-3.70E-04	7.49E-06	-49.45	0	169	0.87
-2.17E-04	8.23E-07	-263.39	0	171	0.97
-2.06E-04	1.09E-06	-189.91	0	175	0.98
-2.22E-04	3.00E-07	-739.19	0	181	0.72
-2.21E-04	2.32E-06	-95.43	0	183	0.57
-2.11E-04	1.83E-06	-115.51	0	187	0.88
-2.34E-04	3.28E-06	-71.32	0	192	0.93
-2.22E-04	1.44E-06	-153.82	0	193	0.63
-2.22E-04	6.81E-07	-325.62	0	194	0.60
-2.01E-04	3.46E-07	-579.79	0	195	0.91

TABLE 4.6: Temperature coefficient results by 3-digit zip code for spatial regression

geographical patterns, figure 4.37 shows the choropleth that was constructed from table 4.6. Again, the temperature dependence seems to be more or less equal throughout the city.

The wind dependence shows a more variable behaviour in different parts of the city. According to table 4.7, values between $3.3 \cdot 10^{-5}$ and $10 \cdot 10^{-5}$ are encountered. Two instances of the coefficient have a negative value, as stated earlier. It may be noticed that, despite the temperature and wind speed variation being of the same order of magnitude, the wind regression coefficients are generally one order of magnitude smaller than those for the temperature.
When the coefficients are plotted on a choropleth map, a result as shown in figure 4.38 is obtained. The larger values are encountered in and around the city center, whereas smaller values can be found north of the city and in other specific locations. This seems

Coeff.	Std. Err.	t-value	P-value	Zip3	R Squared
9.91E-05	7.94E-06	12.49	0	111	0.30
8.21E-05	1.53E-06	53.70	0	112	0.73
8.85E-05	8.31E-07	106.49	0	113	0.89
9.42E-05	2.51E-06	37.54	0	114	0.53
7.78E-05	2.34E-06	33.24	0	115	0.66
8.78E-05	8.87E-07	99.05	0	116	0.92
8.01E-05	1.53E-06	52.42	0	117	0.86
8.76E-05	1.25E-06	70.00	0	118	0.87
4.83E-05	2.54E-06	19.00	0	120	0.74
7.41E-05	1.97E-06	37.55	0	121	0.71
5.69E-05	1.56E-06	36.34	0	122	0.81
6.84E-05	3.46E-06	19.79	0	123	0.66
6.37E-05	1.02E-05	6.23	4.78E-10	124	0.16
6.78E-05	2.66E-06	25.45	0	125	0.75
8.18E-05	2.89E-06	28.28	0	126	0.64
6.02E-05	6.69E-06	9.00	0	127	0.29
4.70E-05	1.99E-06	23.58	0	128	0.82
7.14E-05	7.11E-06	10.04	0	129	0.23
6.33E-05	6.37E-06	9.94	0	131	0.48
8.07E-05	2.52E-05	3.20	0.001353999	132	0.13
4.22E-05	2.71E-06	15.57	0	162	0.76
5.43E-05	2.38E-06	22.78	0	163	0.86
7.09E-05	2.87E-06	24.76	0	164	0.83
5.74E-05	2.94E-06	19.53	0	165	0.62
6.89E-05	2.12E-06	32.47	0	167	0.81
3.70E-05	1.74E-06	21.28	0	168	0.78
-1.14E-04	4.80E-05	-2.38	0.017740064	169	0.87
8.34E-05	5.27E-06	15.81	0	171	0.97
6.32E-05	6.97E-06	9.08	0	175	0.98
8.19E-05	1.93E-06	42.55	0	181	0.72
3.63E-05	1.49E-05	2.44	0.014803994	183	0.57
-1.30E-05	1.17E-05	-1.11	0.265485532	187	0.88
8.66E-05	2.10E-05	4.13	4.60E-05	192	0.93
3.27E-05	9.24E-06	3.54	4.09E-04	193	0.63
3.88E-05	4.36E-06	8.89	0	194	0.60
4.27E-05	2.22E-06	19.20	0	195	0.91

TABLE 4.7: Wind speed coefficient results by 3-digit zip code for spatial regression

to be contrary to the intuition that buildings in densely built areas are more shielded from wind than those in less densely built areas.

Next value of interest is the coefficient of multiple determination, or R squared as indicated in the table: it indicates the proportion of the variation in EUI* that can be explained by variation in the independent variables. In this case, the higher the value, the better the fit of the regression model. Remark that this value is not related to the separate coefficients, but is a characteristic of the regression model. Consequently, its value is repeated for each zip code in the three tables. Further, remark that it is normal practice to adjust this value for the use of multiple regressors, because it will always be higher when an additional explanatory variable is added to the model. This adjustment

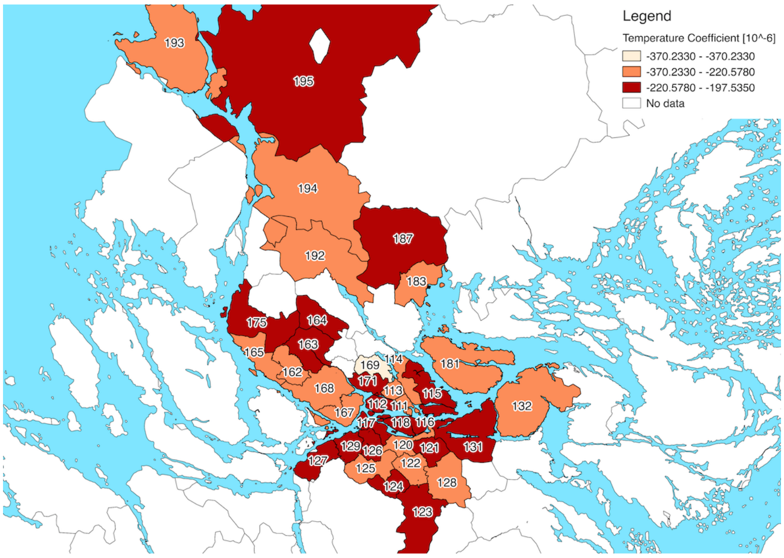


FIGURE 4.37: Choropleth map of the temperature coefficient for spatial regression

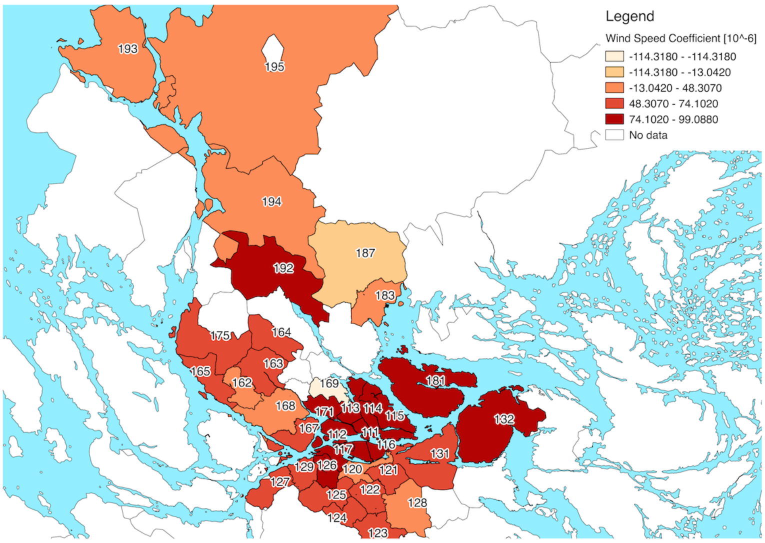


FIGURE 4.38: Choropleth map of the wind speed coefficient for spatial regression

is implemented by the following formula (Navidi, 2008):

$$\bar{R}^2 = R^2 - \left(\frac{k}{n-k-1}\right) R^2, \tag{4.2}$$

with $\bar{R}^2$ the adjusted R^2, k the number of regressors and n the sample size of the regression input. In most cases that are encountered in these regression analyses, n in the denominator is of such a magnitude that the second term in this equation can be safely neglected, and thus the unadjusted R^2 is used.

Apparently, the highest R squared value among the zip codes occurs in 175, where its value is 0.98, and the lowest value is found in 132 with a value of 0.13. The use of this parameter is not clear from the table, therefore it is visualised in another choropleth (figure 4.39).

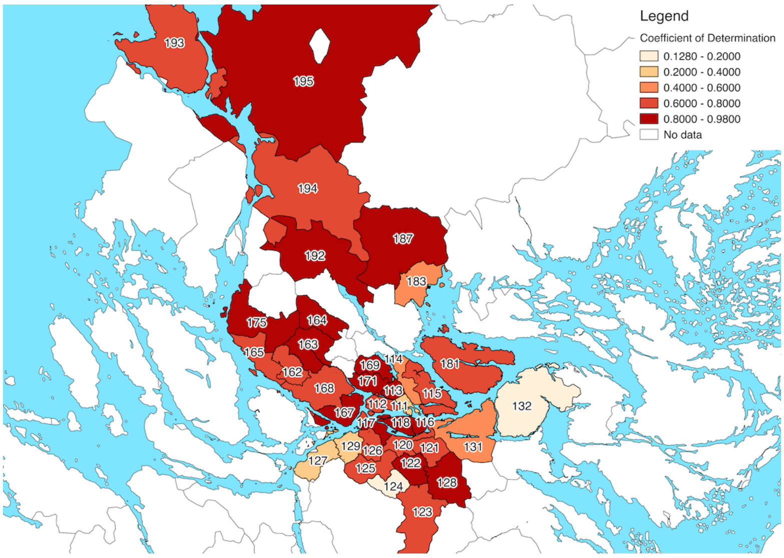


FIGURE 4.39: Choropleth map of the coefficient of multiple determination (R squared) for spatial regression

4.7.2 Spatial regression grouped by consumption

In order to check for different weather dependence for residential buildings with different consumption intensities, the building stock was divided over four groups. As described in paragraph 3.8.2, these four groups can be characterised as buildings with highest, high, medium and low consumption. In order not to overload the report too much with tables, the comprehensive regression results are included in the tables in appendix B. Only the most important results are treated in the current section.

A quick look through the resulting tables in appendix B already tells a lot about the quality of the regression. For most of the zip codes and coefficients, the P values are equally low as in the previous section (none above the 0.05 required for the significance level of 5%), while the R squared values are in general high to very high. Some attention must be paid here: since the regression becomes more and more split up, the number of observations becomes lower, which could imply that the assumption that $R^2 \approx \bar{R}^2$ might not be as accurate as before.
However, for the coefficients in the low consumption group, the R squared values are generally lower. While the P values for the intercept and temperature relation are still 0, now more than half of the zip codes have a wind speed coefficient with a P value that approaches or surpasses 0.05. This indicates that the results for the low consumption class are less accurate and should not be used to draw conclusions.

In order to visualise the difference in weather dependence for the four classes, they have been plotted in the figures below. Figure 4.40 shows the change of the EUI* intercept with respect to the zip code, and grouped per consumption class. Strikingly, the intercept of the low consumption group is in general the highest. However, as remarked in the previous paragraph, the low consumption group displays rather low coefficients of determination. The intercepts for the medium and high consumers seem to be more or less equal, but slightly higher for the medium group. The intercept for the highest consumption group is highly variable, but is often found near the intercept of the medium and high consumers.

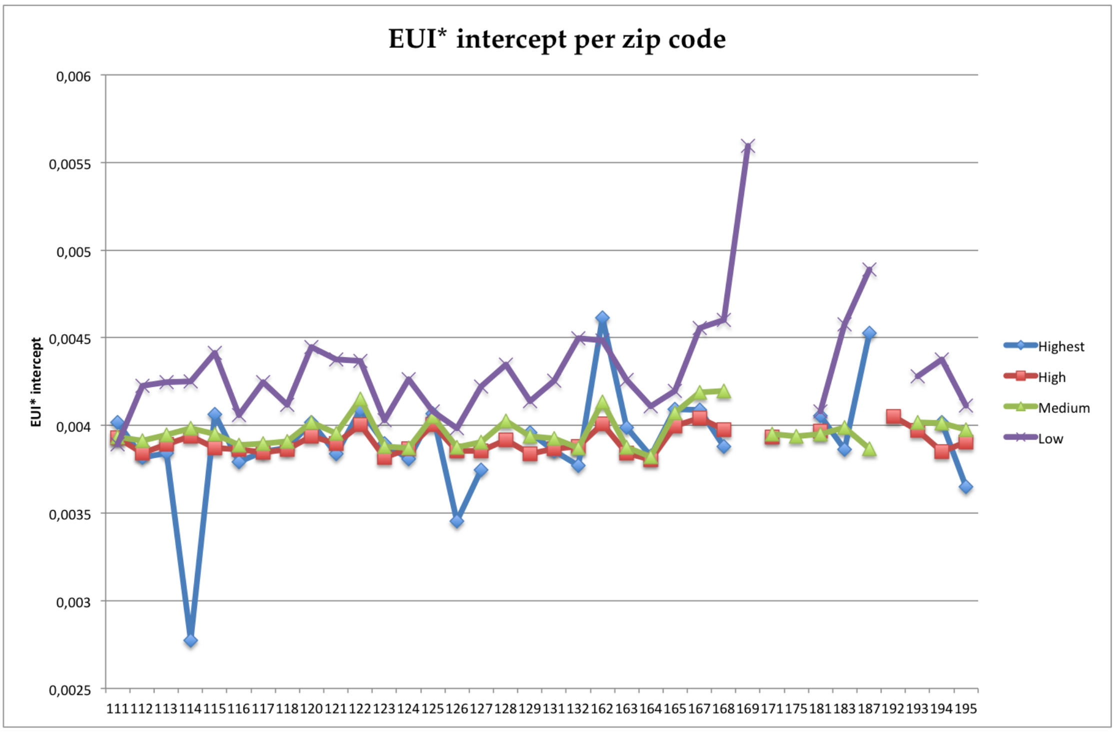


FIGURE 4.40: Comparison of EUI* intercepts for different consumption groups per zip code

The comparison of the temperature and wind dependencies is plotted in figure 4.41 and 4.42 respectively. The temperature dependence graph shows the absolute value (negative) of the temperature regression coefficient. Thus, a higher value on the graph indicates a higher dependency.

Apparently, interdependence of temperature and EUI* is the highest for the low consumption group. This result must be interpreted carefully, since the R squared values for low consumption buildings are low. Consequently, the linear relation for medium and high consumption buildings is increasingly lower. Again, the course of the coefficients for the highest consumption group has a higher variability. Sometimes, buildings that have a high consumption intensity are depending more on the outside temperature, but in general, the dependence becomes less important with higher EUI.

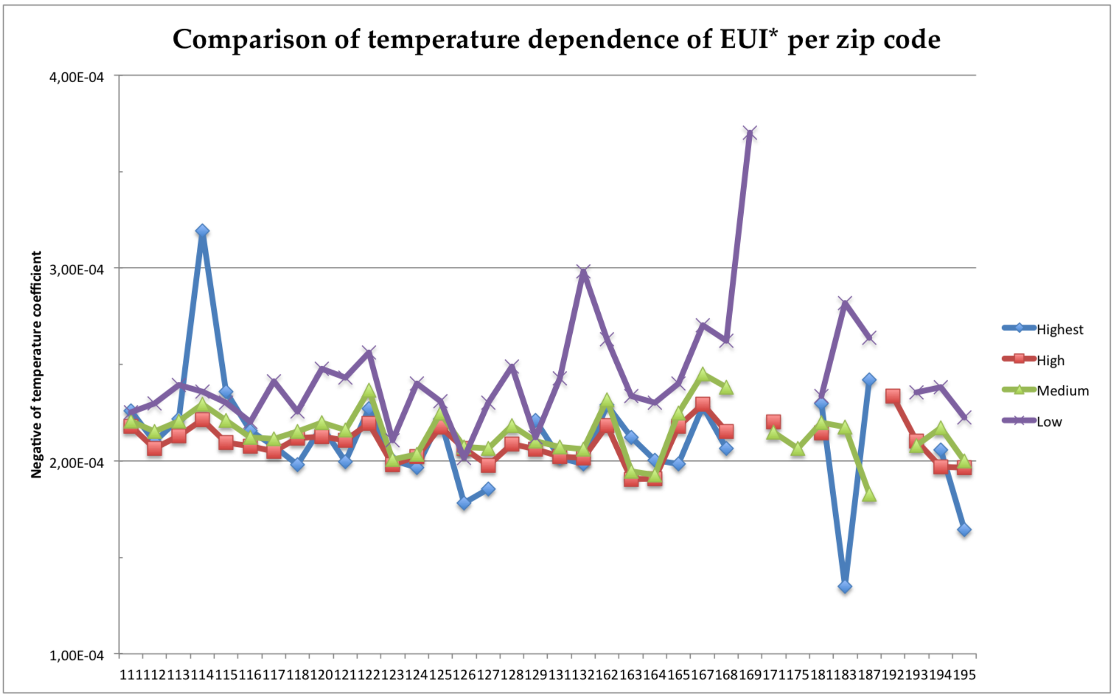


FIGURE 4.41: Comparison of temperature influence for different consumption groups per zip code

The scale of the vertical axis for the wind speed plot (figure 4.42) is chosen with a focus on the stabler medium and high consumption group coefficients. The extremely high and low values in the two other consumption groups are esteemed less important for the sake of the analysis.
Contrary to the temperature dependence, the wind speed dependence is slightly less for the medium consumption group than for the high consumers. The course of the coefficients for the highest consumption group is inconclusive – sometimes higher than the low and medium group, sometimes lower. Regardless of the trustworthiness of the low consumption coefficients, they seem to be generally lower than the other coefficients. This contrasting behaviour for the two coefficients must be investigated closer.

For the temperature and wind coefficients, a slightly decreasing trend from the low zip codes to the higher numbers is apparent. Since the lower postal numbers are located in the city center, this again indicates a higher weather influence in the city center than in the suburbs.

4.7.3 Regression grouped by building category

In a final regression analysis, the spatial aspect is briefly omitted, in order to focus on weather influences in different building types. As seen in table 3.1, there are five categories within the residential building class, viz.

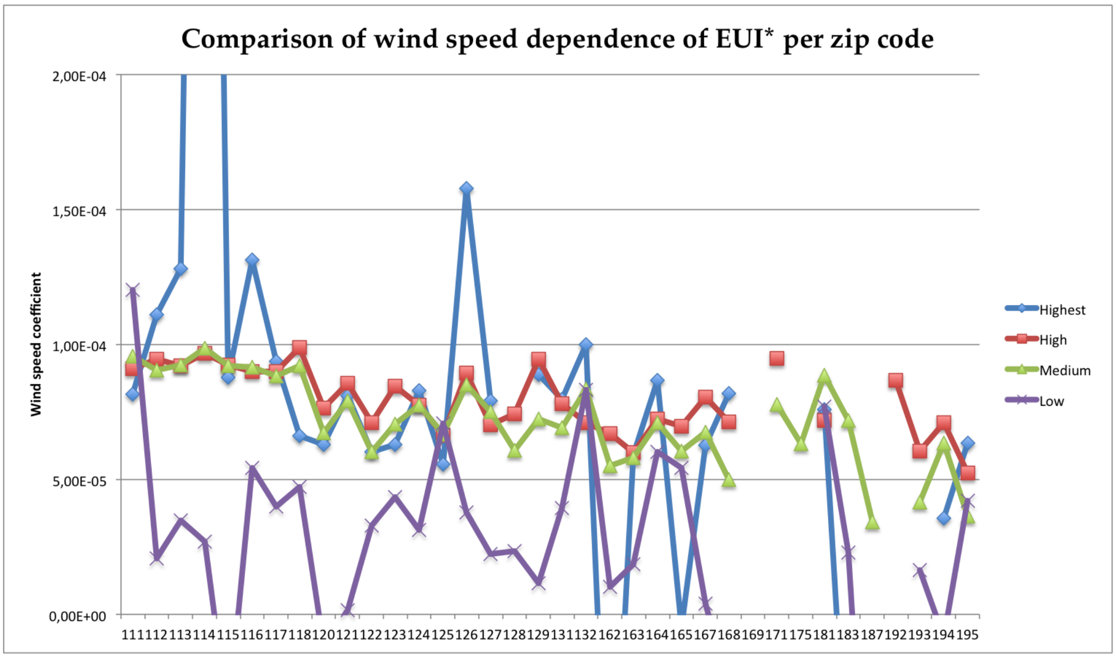


FIGURE 4.42: Comparison of wind speed influence for different consumption groups per zip code

- Single-family & detached house,
- Multi-family house with additional function,
- Apartment in detached multi-family house,
- Detached multi-family house and
- Multi-family house in building block.

Table 4.8 summarizes the results for the regression analysis per residential building group. It can be seen that the intercept is approximately $4 \cdot 10^{-3}$ again, the temperature coefficient lies around $-2.2 \cdot 10^{-4}$ and that the wind speed coefficient fluctuates between $4.55 \cdot 10^{-5}$ and $8.51 \cdot 10^{-5}$. Although the P values indicate undoubtedly that these coefficients are significantly different from 0, the R squared values are on the low sides. This can in part be explained by the high number of observations that have been regressed, bringing along a higher variability and consequently a larger prediction error.

Although the table is not too elaborate, a parallel coordinates chart can help visualising patterns. Such a chart is presented in figure 4.43. The values of the regression coefficients are normalised with respect to the lowest (0) and highest (1) observation. These respective values can be looked up in table 4.8.

In this way, it can be observed that the single-family homes have by far the largest intercept value and temperature influence (the minimal value for the temperature influence

Coefficient	Coeff.	Std. Err.	t-value	P-value	R squared
Detached multi-family house					
Temperature	-2,12E-04	1,61E-07	-1311,97	0,00	0,55
Wind speed	6,83E-05	1,03E-06	66,02	0,00	0,55
Intercept	3,96E-03	3,59E-06	1102,80	0,00	0,55
Multi-family house in building block					
Temperature	-2,17E-04	1,40E-07	-1550,72	0,00	0,59
Wind speed	8,51E-05	8,97E-07	94,94	0,00	0,59
Intercept	3,94E-03	3,11E-06	1266,94	0,00	0,59
Multi-family house with additional function					
Temperature	-2,27E-04	2,37E-07	-956,84	0,00	0,87
Wind speed	7,60E-05	1,52E-06	50,03	0,00	0,87
Intercept	4,04E-03	5,27E-06	766,09	0,00	0,87
Apartment in detached multi-family house					
Temperature	-2,05E-04	1,15E-06	-178,17	0,00	0,36
Wind speed	4,55E-05	7,38E-06	6,17	0,00	0,36
Intercept	3,98E-03	2,56E-05	155,59	0,00	0,36
Single-family & detached house					
Temperature	-2,44E-04	2,04E-07	-1195,59	0,00	0,62
Wind speed	5,42E-05	1,31E-06	41,48	0,00	0,62
Intercept	4,22E-03	4,54E-06	930,26	0,00	0,62

TABLE 4.8: Results of regression grouped by building category

corresponds with the highest influence), but a low influence of wind speed. The multi-family homes with additional function are positioned in the intermediate region for all coefficients. The three remaining building types have more or less the same intercept value and display the same low temperature influence. However the wind influence is very different for these three types: while the apartments inside a larger building experience the smallest effect from wind, the detached multi-family homes and multi-family homes in a block of buildings are affected increasingly.

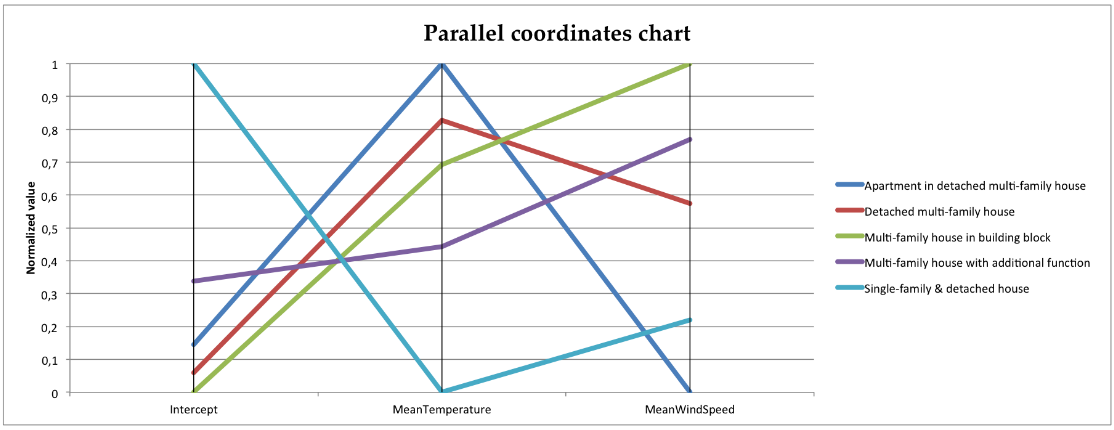


FIGURE 4.43: Parallel coordinate chart of the regression coefficients per building type

4.7.4 *Key findings*

- There is sufficient statistical evidence from the studied data sets to assume that heating consumption in buildings in the city center have a higher dependence on the wind speed than buildings in the periphery of Stockholm.
- The influence of temperature is the largest, though, and remains more or less constant across Stockholm.
- Buildings that have a higher energy intensity (EUI) display an overall higher influence of wind speed on energy consumption. The case for temperature influence remains inconclusive.
- There are clear differences between the influence of weather circumstances on energy consumption for different categories of residential buildings.

Chapter 5

Discussion and conclusions

5.1 Discussion

5.1.1 Usefulness of clusters

Clusters can be a useful tool to investigate which buildings have the highest savings potential and should be prioritised in a retrofitting plan. However, the clusters that were developed in this thesis do not differ greatly in their average time of consumption. A possible explanation for this could be that the number of clusters is not high enough, but it could also be caused by a more uniform distribution of district heating consumption. While the consumption of electricity (as in the white paper by Silipo and Winters (2013)) clearly depends on the behaviour of the consumer, the consumption data in this research displays a recurring consumption pattern with two peaks per day and a smaller nighttime consumption. This assumption is partly confirmed by the time-dependence analysis (section 4.2).

5.1.2 Reliability of savings scenarios

A number of energy saving scenarios has been defined in this thesis. In the first place, there were the high, medium and low savings scenarios in the total savings potential analysis. As the name already gives away, the purpose is to explore what are the maximum possible savings, without restricting itself by defining how these savings should be realised. In this sense there is no reliability issue for this type of explorative analysis. The second savings calculation that also proposes a set of actions that should be undertaken is more dependent on the plausibility of the chosen variables.

EUI variation from efficiency However, because of the heterogeneity of the buildings (6 classes, 30 categories), it might not be realistic to assume that all buildings can achieve the same EUI goal. For instance, industrial buildings might use the energy input of the DH system for other purposes than heating alone (e.g., process heat). This is also the reason why industrial buildings were not considered in the retrofit plan analysis: the EUIs in this category are much more varied than for other building types, and this variation might not be a result of consumption inefficiency alone.

External influences Further, since the analysis simulates the investment for savings measures and their return, it is highly dependent on the evolution of external parameters (e.g., price of DH, effectiveness and actual cost of savings measures…) as well. The sensitivity analysis estimates the influence of two of these uncertainties.
As it appears, the influence of the most costly retrofitting measure (climate shell renovation) has a major influence on the savings that can be easily achieved. As stated in the results, for the lower price range, the variation on the possible savings is rather large, but the average savings are high, while for the higher price range the savings are low but rather invariable. The presented analysis for the maximum profitable savings uses a price value of 2000 kr/m^2, thus in the lower range. As a consequence, these savings are rather uncertain and more likely to be an overestimation. These results need to be interpreted carefully.
On the other hand, the planning horizon is also of great influence. If the average anticipated pay-back time is to be decreased to 10 years, the decrease in number of investments might lead to a reduction of the savings by half.

Climate influence Although the used dataset is exceptionally large, it only presents data from one year. This means that the projected savings will only be correct if the average annual EUIs are more or less the same. It would be interesting to use data from multiple years as an input, but this poses another difficulty, since the data under different weather circumstances are not directly comparable. Therefore, the EUI values should be weather-normalised, using a technique that can be found generally in papers in this domain of study. It was not applied here, since there was no data from different years to compare.

5.1.3 Use of results by policy makers and replicability

A part of this thesis has investigated a strategy that could be adopted by the City of Stockholm in order to realise the goal of 5% building energy savings for 2012-2015. In addition, the total savings potential if all buildings were to comply with the building regulations from Boverket was calculated. This methodology can help policy makers and governmental institutions to make better decisions and to achieve their goals for energy and climate in the most efficient way possible.

Even more so, this methodology can be generalized and applied in cities all over the world. The only condition to be fulfilled is that consumption data for a large enough set of buildings is available for analysis. Using KNIME, a similar analysis with clusters and annual savings can easily be set up in other cities.

5.1.4 Retrofitting plan

General comment on savings In this analysis, the goal was set at 5% energy savings according to the climate action plan of the city of Stockholm. These savings relate to all building energy consumption, not just heating consumption. While assuming that the same savings would apply for all building groups in Stockholm (i.e. that buildings with DH would save the same proportion as other buildings) is straightforward, the 5% for heating consumption is only correct if the same amount is saved for electricity consumption as well. Indeed, according to Lönngren (2012), heating consumption encompasses 70% of the total consumption, the rest being taken up by electricity.

Comment on benefits In the current analysis, the benefits of retrofitting are only expressed as monetary savings because of the consumption reduction. Tommerup and Svendsen (2006) point out that the benefits should be seen from the perspective of sustainable development as well. Furthermore they mention the fact that additional benefits lie in the improvement of living conditions in the retrofitted buildings and in the rise in property value because of better energy performance.
Another comment is on the assumption of constant district heating prices. On the one hand, an increase of the DH heating cost can be expected from the historically rising prices. On the other hand, the proposed savings measures would impact the current need for peak plants to deliver peak demand. When these peaks are lowered, the total running costs will be lower as well, and from this a negative growth of the DH cost can be expected.

Comment on Stockholm's Climate Action Plan The retrofitting plan analysis has mainly focused on examining the impact of the proposed energy saving measures from the climate action plan (Lönngren, 2012). The largest difference with the assumptions in the report (except for the focus on heating consumption), is that the report only considers residential and commercial (office) buildings. Figure 4.27 shows that this assumption is justified, but nevertheless savings in other types of buildings are also possible. With these results, the city government is encouraged to further assess savings possibilities in buildings that belong to these other classes.

Financing mechanisms Another question related to the retrofitting plan is how to incentivise or finance the savings measures. Since the costs of retrofitting are very high, it is unlikely that private actors will invest on their own initiative. The most straightforward strategy would be to subsidise those properties that are part of the retrofitting plan that was compiled.

On the other hand, a more complicated system would enable even higher savings. Figure 5.1 shows a different representation of the cost of investment, the monetary benefits from energy savings and the percentage of energy that can be saved. Although the saved running costs depend mainly on the chosen planning horizon (this graph was based on the 15 year horizon), it is clear that the "general" break-even point lies at a much higher savings proportion than the positive ROI case. If a system could be invented in which the extra benefits from the buildings with the highest ROI could be invested in less profitable retrofittings, this could mean a considerable increase of the possible savings.

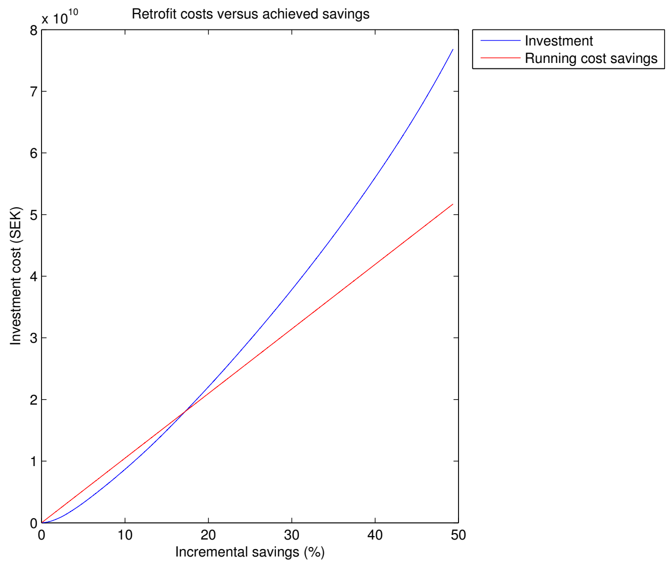


FIGURE 5.1: Monetary savings from saved energy vs. total investment

5.1.5 Weather influence regression analysis

One general remark about these linear regression analyses is that the three criteria for this technique (normal distribution of variables, random and independent errors and constant variance) have not been checked. Further, it could be interesting to investigate the prediction errors for patterns in order to see if a transformation of the independent variables temperature and wind speed would have been needed. This was however not done because of a larger focus on other aspects in the thesis.

Furthermore, it must be said that the current analysis was performed on a data set of one year, and only for residential buildings. In order to yield more general results, the analysis should be performed on a range of multiple years and other building classes could be included for completeness.

Simple spatial regression In the case where the only distinction that is made, are the zip code areas, the results seem reliable, but counterintuitive. It would be expected that buildings in the city center – because of the higher density of buildings – would experience less influence from wind and temperature. This effect is closely related to the urban heat island phenomenon (Dorer et al., 2013). However, the observations show that the influence of the temperature is hardly different, and that the influence of temperature is greater in the city center than in the peripheral areas.
It can be argued that the influence is biased by using the EUI* instead of the actual EUI, or that more information about the buildings' shape and local weather information are needed. Further, if one looks back at figure 4.13, it can easily be seen that the largest variation of the overall consumption is already explained by the variation in temperature. Therefore, the relation with the wind speed is much weaker, which could be an explanation for these counterintuitive results.

Grouped by consumption The observations for this grouped regression analysis are more or less the same as for the non-grouped one. However, the difference of the coefficients across different consumption groups is clear and again poses some interpretation difficulties.
For instance the variation of the intercept: the fact that the EUI* intercept is higher for buildings with a lower average EUI seems odd. The wind speed coefficient should have no influence on the intercept, since the extra consumption from 0 wind speed should be negligible. Reasoning that, independent of the average consumption (since EUI* is dimensionless), the average EUI* value should be one, the difference in intercept would be logical if the temperature coefficient for low consumption would be larger than for high consumption. As a matter of fact, this is the case (see figure 4.42), which explains the strange intercept behaviour.
On the other hand, it is not entirely clear why the temperature coefficient is higher for buildings with low consumption. A possible explanation is that the consumption in higher consumption groups is high irrespective of the outside temperature, or that the use of the dimensionless EUI* mitigates the consumption difference and thus makes the influence smaller.
Anyway, these results must be interpreted carefully; the reason for this is the lower R squared value for the analysis grouped by consumption.

Grouped per building category The reasoning in the previous paragraph seems to work for the analysis per building category as well: the building groups with lower intercepts tend to have a lower temperature influence as well, whereas the wind speed coefficient is not related.
Further, it seems that building that are more encapsulated by other buildings (apartments, buildings in blocks) have a lower influence from temperature. The same reasoning is *not* observed in the wind speed coefficient: although single apartments display the lowest wind dependence, the multi-family homes in a building block have a higher wind influence. The single-family homes (which are expected to have the highest surface/volume ratio and hence the highest convection losses) also have a smaller wind influence, in contrast with the expectation.
Again, the remark about the smaller statistical significance of these results applies.

5.2 Conclusions

This thesis has pointed out that there is considerable room for improvement with respect to energy use intensity among the district heating customers of the district heating network.

Referring to the objectives, it can be concluded that the goals that were set have been achieved. Understanding of the input data set was provided in the descriptive statistics section, in which the contributions of different types of customers to the total consumption were analysed. Residential buildings constitute the largest consumer group, followed by commercial buildings. In terms of energy efficiency, commercial and health & care buildings perform best. Since the data set was in an exceptionally good shape, no more curation than the removal of occasional outliers was needed. For ease of analysis, the building categories have been combined into six building classes.
KNIME has been used for most of the data processing steps, but during the analysis process, additional software packages were introduced because they were more appropriate for some tasks. The clustering analysis has not provided the anticipated results, but proved a useful tool for the purpose of targetting inefficient user groups. Slight evidence was found for the influence of location and weather on energy and consumption; however, the actual building characteristics (vintage, category, size) appear to have a much larger influence on the total consumption and energy use intensity.
Furthermore, most of the focus was on the identification of savings potential. Two separate analyses were set up, one set of hypothetical scenarios that investigated the maximum potential for energy efficiency improvement, another set of more realistic retrofitting measures with an economic analysis. The maximal potential for energy savings in the studied set of buildings was estimated at 59% (4.6 TWh), while the maximal economically feasible savings are limited to ca. 6% (472 GWh).

Chapter 6

Recommendations

To researchers

In order to get a better understanding of the counter-intuitive results for the weather influence analysis, a larger study with more years' worth of data (both energy consumption and weather data) would be needed. On the other hand, more accurate data from a number of pilot buildings – instead of analyzing circa 11 000 residential buildings – could yield more significant results. One way or another, more information about the buildings' shape, effective (shell) surface, heat losses and average energy consumption would make a much stronger case, since a higher focus on the effect of the weather without secondary effects would be possible.

Since a lot of uncertainty exists about what types of savings would be possible and how to model the effect and cost of energy efficiency measures, it would be very interesting to perform a pilot study on multiple types of houses (representative for a larger proportion of a building stock). The resulting parameters such as cost, lowest possible EUI, proportion of energy saved... could serve as a much more precise input for this kind of energy saving simulations.

In general, it would be interesting to compare energy consumption data from this year to the consumption of the same building set in earlier years. This would again improve the results of the energy saving analysis. Furthermore, comparison of weahter-normalised consumption data over the years can reveal cases of performance degradation in different types of buildings. Finally, the addition of electricity data and a corresponding savings analysis is recommended.

To local governments and policy makers

The policy makers in other (smart) cities are encouraged to gather energy data with high time-resolution and make it available for research. In addition to consumption data, it is encouraged to provide information about the buildings' location, area, shape and type of heating used. Nice to haves for more specific analysis are information about energy efficiency measures already installed, types of fenestration, number of inhabitants...

Further, it is recommended that policy makers thoroughly study the expected performance of savings measures that they propose. The method that was used in this thesis is a very strong tool to analyse the results of regulations, while at the same time the buildings with the best return on investment can be separated from the rest.

To the City of Stockholm

The City of Stockholm is encouraged to use the results of the retrofitting analysis in order to refine the climate and energy action plan. This thesis has shown that energy savings cannot only be achieved in residential and commercial buildings. Furthermore, it is clear that there is a clear distinction between buildings as to cost efficiency of investments.

These findings only concern a limited set of buildings that use district heating. The City can extend this analysis with more buildings from Stockholm, and it is highly encouraged to collect consumption information with high time resolution from these buildings as well. This thesis has proved that the 5% energy goal (2012-2015) that was set by the Stockholm government is feasible at reasonable cost, and with pay-back of the initial investments (with the used calculation parameters).

Key recommendation Instead of randomly retrofitting buildings to increase consumption efficiency, the local government should investigate which types of buildings yield the best results after retrofitting (using clusters) and locate areas in the city with high inefficiencies (using savings maps). Only in this way will the city's energy saving goals be reached in an economically sustainable way.

Bibliography

Abu-Mostafa, Y. (2012), *Learning from Data - Online Machine Learning Course*, California Institute of Technology (Caltech), Pasadena, CA.
URL: *http://work.caltech.edu/telecourse.html*

Boverket (2011), 'Karta'.
URL: *http://www.boverket.se/Om-Boverket/Nyhetsbrev/Boverket-informerar/Ar-2011/20113/*

Boverket (2013), 'Boverkets byggregler - föreskrifter och allmänna råd, BBR'.
URL: *http://www.boverket.se/Global/Lag_o_ratt/Dokument/Boverkets-Forfattningssamling/BBR-konsoliderad-BFS2011-6-tom-BFS2013-14.pdf*

Caragliu, A., Del Bo, C. and Nijkamp, P. (2011), 'Smart Cities in Europe', *Journal of Urban Technology* **18**(2), 65–82.
URL: *http://www.tandfonline.com/doi/abs/10.1080/10630732.2011.601117*

Cukier, K. and Mayer-Schoenberger, V. (2013), 'The Rise of Big Data.', *Foreign Affairs* **92**(3), 27–40.

Danielski, I. (2012), 'Large variations in specific final energy use in Swedish apartment buildings: Causes and solutions', *Energy and Buildings* **49**, 276–285.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0378778812000928*

Dorer, V., Allegrini, J., Orehounig, K., Moonen, P., Upadhyay, G., Kämpf, J. and Carmeliet, J. (2013), 'Modelling the Urban Microclimate and its Impact on the Energy Demand of Buildings and Building Clusters', *Proceedings of BS2013* pp. 3483–3489.
URL: *http://infoscience.epfl.ch/record/195745*

Ehrenfeld, J. (2004), 'Industrial ecology: a new field or only a metaphor?', *Journal of Cleaner Production* **12**(8-10), 825–831.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S095965260400068X*

Fortum (2014), Redovisning av kostnader för fjärrvärme enligt Energimarknadsinspektionens krav, EIFS 2009:3, Technical report, Stockholm.

URL: *http://www.fortum.com/countries/se/SiteCollectionDocuments/kostnadsredovisning-fjarrvarme-2014-foretag.pdf*

Graedel, T. and Allenby, B. (2010), *Industrial Ecology and Sustainable Engineering*, 2 edn, Prentice Hall, Upper Saddle River, NJ.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *The elements of statistical learning*, second edn, Springer, New York, NY.
URL: *http://www-stat.stanford.edu/ tibs/book/*

Hey, T., Tansley, S. and Tolle, K., eds (2009), *The Fourth Paradigm: Data-intesive scientific discovery*, Microsoft Research, Redmond, Washington.
URL: *http://research.microsoft.com/en-us/collaboration/fourthparadigm/*

Hollands, R. G. (2008), 'Will the real smart city please stand up?', *City* **12**(3), 303–320.
URL: *http://www.tandfonline.com/doi/abs/10.1080/13604810802479126*

IEA (2012), *World Energy Outlook 2012*, International Energy Agency, Paris.
URL: *http://www.iea.org/publications/freepublications/publication/name,49561,en.html*

IPCC (2014), *Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge.
URL: *http://www.ipcc.ch/report/ar5/*

Jenks, G. F. and Caspall, F. C. (1971), 'Error on choroplethic maps: definition, measurement, reduction', *Annals of the Association of American Geographers* **61**(2), 217–244.
URL: *http://www.tandfonline.com/doi/abs/10.1111/j.1467-8306.1971.tb00779.x*

Lazaroiu, G. C. and Roscia, M. (2012), 'Definition methodology for the smart cities model', *Energy* **47**(1), 326–332.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0360544212007062*

Lienhard IV, J. H. and Lienhard V, J. H. (2008), *A heat transfer handbook*, third edn, Phlogiston Press, Cambridge, MA.
URL: *http://web.mit.edu/lienhard/www/ahttv131.pdf*

Lloyd, C. D. (2010), *Spatial data analysis: an introduction for GIS users*, Oxford : Oxford University Press, Oxford.

Lönngren, O. (2012), Stockholms åtgärdsplan för klimat och energi 2012-2015 med utblick till 2030, Technical report, Stockholms Stad, Stockholm.
URL: *http://www.stockholm.se/OmStockholm/Stadens-klimatarbete/Atgarder-som-lett-till-betydande-minskningar-av-utslappen/Stockholms-atgardsplan-for-klimat-och-energi/*

MacKay, D. J. C. (2003), *Information theory, inference and learning algorithms*, fourth edn, Cambridge University Press, Cambridge.
URL: *http://www.inference.phy.cam.ac.uk/itprnn/book.pdf*

Magnusson, D. (2013), District Heating in a Liberalized Energy Market: A New Order? - Planning and Development in the Stockholm Region, 1978-2012, Phd thesis, Linköping University.
URL: *http://www.diva-portal.org/smash/record.jsf?pid=diva2:605579*

McGill, R., Tukey, J. and Larsen, W. (1978), 'Variations of box plots', *The American Statistician* **32**(1), 12–16.
URL: *http://amstat.tandfonline.com/doi/abs/10.1080/00031305.1978.10479236*

Meadows, D. H., Meadows, D. L., Randers, J. r. and Behrens, W. W. (1972), *The limits to growth*, Universe Books, New York.
URL: *http://www.donellameadows.org/wp-content/userfiles/Limits-to-Growth-digital-scan-version.pdf*

Moran, M. J., Shapiro, H. N., Boettner, D. D. and Bailey, M. B. (2011), *Fundamentals of Engineering Thermodynamics*, 7 edn, Wiley, Hoboken, NJ.

Nässén, J. and Holmberg, J. (2005), 'Energy efficiency—a forgotten goal in the Swedish building sector?', *Energy Policy* **33**(8), 1037–1051.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0301421503003355*

Navidi, W. C. (2008), *Statistics for engineers and scientists*, third edn, McGraw Hill, New York, NY.

Peterson, K. and Crowther, H. (2010), 'Building energy use intesity', *High Performing Buildings Magazine* pp. 40–50.
URL: *http://www.hpbmagazine.org/File Library/Case Studies/Summer 2010/EUI.pdf*

Silipo, R. and Winters, P. (2013), Big Data , Smart Energy , and Predictive Analytics Time Series Prediction of Smart Energy Data, Technical report, KNIME, Konstanz.

Sjögren, J.-U. (2007), 'Användning av kall- och varmvatten i flerbostadshus', *Energi & miljö* (11).

Sveriges Meteorologiska och Hydrologiska Institut (2014), 'SMHI Öppna data'.
URL: *http://opendata-download-metobs.smhi.se/explore/#*

Tibshirani, R., Walther, G. and Hastie, T. (2001), 'Estimating the number of clusters in a data set via the gap statistic', *Journal of the Royal Statistical Society* .
URL: *http://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00293/abstract*

Tommerup, H. and Svendsen, S. (2006), 'Energy savings in Danish residential building stock', *Energy and Buildings* **38**(6), 618–626.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0378778805001969*

Touchie, M. F., Binkley, C. and Pressnail, K. D. (2013), 'Correlating energy consumption with multi-unit residential building characteristics in the city of Toronto', *Energy and Buildings* **66**, 648–656.
URL: *http://linkinghub.elsevier.com/retrieve/pii/S0378778813004593*

United Nations (2012), 'World Urbanization Prospects: The 2011 Revision'.
URL: *http://esa.un.org/unpd/wup/CD-ROM/Urban-Rural-Population.htm*

US Census Bureau (2014), 'Population Clock'.
URL: *http://www.census.gov/popclock/*

Waggoner, P. and Ausubel, J. (2002), 'A framework for sustainability science: A renovated IPAT identity', *Proceedings of the National Academy of Sciences of the United States of America* .
URL: *http://www.pnas.org/content/99/12/7860.short*

Webb, A. R. and Copsey, K. D. (2011), Clustering, *in* 'Statistical Pattern Recognition', John Wiley & Sons, Ltd, chapter 11, pp. 501–554.
URL: *http://dx.doi.org/10.1002/9781119952954.ch11*

Weston.pace (2007), *K-means example step 1 to 4*, Wikimedia Commons.
URL: *http://commons.wikimedia.org/wiki/File:K_Means_Example_Step_1.svg*
http://commons.wikimedia.org/wiki/File:K_Means_Example_Step_2.svg
http://commons.wikimedia.org/wiki/File:K_Means_Example_Step_3.svg
http://commons.wikimedia.org/wiki/File:K_Means_Example_Step_4.svg

Wheeler, D. C. and Paéz, A. (2010), Geographically Weighted Regression, *in* M. M. Fischer and A. Getis, eds, 'Handbook of Applied Spatial Analysis', Springer-Verlag Berlin Heidelberg, Heidelberg, chapter C.5, pp. 461–486.

WHO (2013), 'Urban population growth'.
URL: *http://www.who.int/gho/urban_health/situation_trends/urban_population_growth_text/en/*

Appendix A

Descriptive statistics

This appendix lists the averages, minima, maxima and standard variations for the various building categories, grouped per class. These results are referred to in section 4.1.

	Detached multi-family house	Multi-family house in block of buildings	Multi-family house with add. function	Apartment in detached multi-family house	Single-family & detached house
	Annual energy consumption [MWh]				
Mean	766,88	453,86	635,60	8,63	58,82
Std. dev.	1049,01	431,93	778,77	7,47	276,97
Min	0,12	0,00	39,52	0,43	0,03
Max	20187,92	8060,82	6653,10	58,33	6424,83
	Annual energy use intensity [MWh/m^2]				
Mean	0,151	0,142	0,137	0,115	0,117
Std. dev.	7,97E-02	4,05E-02	7,58E-02	5,27E-02	4,79E-02
Min	6,56E-05	6,62E-07	4,86E-03	6,87E-03	2,00E-04
Max	1,989	0,785	1,258	0,285	0,545
	Building area [m^2]				
Mean	5797,32	3391,04	5701,04	90,11	436,97
Std. dev.	7859,36	3236,37	8782,90	160,19	1860,80
Min	100	140	400	33	65
Max	170550	41600	82255	1523	39000
Count	3816	4529	372	155	2454

TABLE A.1: Descriptive statistics of residential meters

	Hotel	Office & shop building	Office building with add. function	Warehouse with office
	Annual energy consumption [MWh]			
Mean	1230,15	907,29	686,29	500,70
Std. dev.	1404,13	1247,31	667,92	618,35
Min	32,61	1,93	2,19	22,74
Max	10170,02	13014,08	4011,61	5570,54
	Annual energy use intensity [MWh/m^2]			
Mean	0,133	0,104	0,115	0,109
Std. dev.	7,66E-02	5,42E-02	7,81E-02	1,28E-01
Min	3,99E-03	1,45E-03	1,75E-03	4,31E-03
Max	0,58	0,43	1,09	1,81
	Building area [m^2]			
Mean	11691	11195	8053	5840
Std. dev.	13799	16898	9854	6047
Min	183	209	284	149
Max	94000	169000	71883	41382
Count	77	1043	331	228

TABLE A.2: Descriptive statistics of commercial meters

	Daycare & recreation center	Service flat	Hospital and nursing home
	Annual energy consumption [MWh]		
Mean	125,63	1146,55	1984,766
Std. dev.	172,39	1045,85	3815,99
Min	2,21	33,28	46,32
Max	1779,36	4114,54	26280,76
	Annual energy use intensity [MWh/m^2]		
Mean	0,147	0,134	0,121
Std. dev.	0,0616	0,0743	0,0503
Min	0,0022	0,01306	0,0143
Max	0,424	0,689	0,297
	Building area [m^2]		
Mean	885	9534	19499
Std. dev.	976	8361	43621
Min	172	198	423
Max	10000	35072	359798
Count	131	102	134

TABLE A.3: Descriptive statistics of health and care meters

	Church	Community & meeting center	School	Sports facility
	Annual energy consumption [MWh]			
Mean	277,11	1188,42	885,50	1463,49
Std. dev.	184,01	2089,54	777,07	4556,82
Min	45,20	12,61	11,36	8,25
Max	804,05	14538,96	5445,12	32750,23
	Annual energy use intensity [MWh/m^2]			
Mean	0,163	0,154	0,133	0,196
Std. dev.	6,33E-02	1,23E-01	6,28E-02	2,06E-01
Min	6,57E-02	3,88E-02	6,25E-04	9,34E-03
Max	0,383	1,055	0,673	1,670
	Building area [m^2]			
Mean	1858	9554	9391	12193
Std. dev.	1319	15406	18353	33263
Min	270	80	120	86
Max	7120	77500	154184	168184
Count	62	73	367	106

TABLE A.4: Descriptive statistics of public meters

	Vehicle assembly	Garage	Printing industry	Workshop	Other facilities	Other industrial facilities
	Annual energy consumption [MWh]					
Mean	1198,34	558,13	2370,32	549,23	1828,58	1184,75
Std. dev.	3449,39	806,08	5059,90	772,59	8557,14	3013,38
Min	4,69	13,31	132,59	14,11	0,20	6,59
Max	23057,63	3307,23	15756,39	6313,62	82946,24	25516,62
	Annual energy use intensity [MWh/m^2]					
Mean	0,172	0,356	0,123	0,133	0,186	0,293
Std. dev.	0,106	1,455	0,030	0,080	0,198	1,339
Min	2,34E-02	4,42E-03	6,88E-02	2,53E-03	1,81E-04	4,85E-03
Max	0,563	8,046	0,168	0,439	1,542	12,002
	Building area [m^2]					
Mean	8426	7194	16220	5298	14316	10037
Std. dev.	25918	6999	29808	8073	48680	20054
Min	100	200	1200	195	1	305
Max	170000	30000	93700	59331	400000	154184
Count	50	30	9	97	153	79

TABLE A.5: Descriptive statistics for industrial meters

	Only warm water	Only heating	T-bana stations	District heating plant, charged	District heating plant, not charged	Street heat	No classification
	Annual energy consumption [MWh]						
Mean	142,63	1105,11	172,93	105,07	836,60	702,99	248,83
Std. dev.	146,45	1388,77	246,89	9,97	1038,77	910,30	296,13
Min	13,14	113,49	11,07	98,02	13,54	12,87	22,86
Max	633,80	3446,20	1247,56	112,12	2576,94	4250,46	990,64
	Annual energy use intensity [MWh/m^2]						
Mean	0,041	0,100	0,148	0,081	0,301	0,085	0,134
Std. dev.	2,18E-02	5,21E-02	1,64E-01	5,83E-02	2,74E-01	6,14E-02	9,89E-02
Min	0,012	0,030	0,006	0,040	0,023	1,61E-03	3,67E-02
Max	0,142	0,170	0,827	0,123	0,715	0,179	0,398
	Building area [m^2]						
Mean	4331	10674	1428	1800	4128	11698	2757
Std. dev.	5672	8639	1404	1414	7547	11101	3295
Min	669	1126	70	800	370	400	139
Max	27600	20250	9071	2800	19400	34650	11310
Count	39	5	87	2	6	21	11

TABLE A.6: Descriptive statistics of other meters

Appendix B

Spatial regression results

This appendix contains a comprehensive list of results for the spatial regression grouped by consumption.

Zip	Coeff.	Std. Error	t-value	P value	R square
111	4,02E-03	3,08E-05	130,38	0,00	0,91
112	3,82E-03	8,53E-05	44,75	0,00	0,73
113	3,84E-03	6,70E-05	57,34	0,00	0,83
114	2,78E-03	0,00197934	1,40	0,16	0,03
115	4,06E-03	2,23E-05	182,17	0,00	0,97
116	3,79E-03	5,76E-05	65,79	0,00	0,95
117	3,85E-03	3,97E-05	97,06	0,00	0,88
118	3,87E-03	4,43E-05	87,42	0,00	0,93
120	4,02E-03	3,66E-05	109,72	0,00	0,92
121	3,84E-03	1,48E-05	258,44	0,00	0,95
122	4,09E-03	7,32E-05	55,83	0,00	0,73
123	3,90E-03	3,77E-05	103,44	0,00	0,95
124	3,81E-03	2,07E-05	184,14	0,00	0,95
125	4,07E-03	2,87E-05	141,90	0,00	0,92
126	3,45E-03	7,19E-05	48,07	0,00	0,58
127	3,75E-03	5,27E-05	71,12	0,00	0,81
129	3,96E-03	4,03E-05	98,33	0,00	0,89
131	3,85E-03	2,12E-05	181,91	0,00	0,96
132	3,77E-03	4,57E-05	82,60	0,00	0,96
162	4,62E-03	1,31E-04	35,34	0,00	0,68
163	3,99E-03	3,25E-05	122,78	0,00	0,97
164	3,83E-03	3,20E-05	119,69	0,00	0,90
165	4,09E-03	5,10E-05	80,20	0,00	0,95
167	4,09E-03	1,84E-05	221,84	0,00	0,91
168	3,88E-03	3,63E-05	106,82	0,00	0,71
181	4,06E-03	3,61E-05	112,32	0,00	0,90
183	3,87E-03	9,70E-05	39,87	0,00	0,47
187	4,53E-03	9,07E-05	49,88	0,00	0,91
194	4,02E-03	6,48E-05	61,97	0,00	0,82
195	3,65E-03	2,29E-05	159,38	0,00	0,96

TABLE B.1: Intercept for highest consumption group

Zip	Coeff.	Std. Error	t-value	P value	R square
111	-2,26E-04	1,38E-06	-163,30	0,00	0,91
112	-2,10E-04	3,84E-06	-54,73	0,00	0,73
113	-2,22E-04	3,01E-06	-73,61	0,00	0,83
114	-3,19E-04	8,90E-05	-3,59	0,00	0,03
115	-2,36E-04	1,00E-06	-235,36	0,00	0,97
116	-2,16E-04	2,59E-06	-83,17	0,00	0,95
117	-2,08E-04	1,78E-06	-116,40	0,00	0,88
118	-1,98E-04	1,99E-06	-99,40	0,00	0,93
120	-2,18E-04	1,65E-06	-132,46	0,00	0,92
121	-1,99E-04	6,67E-07	-298,86	0,00	0,95
122	-2,27E-04	3,29E-06	-69,03	0,00	0,73
123	-2,00E-04	1,69E-06	-118,27	0,00	0,95
124	-1,96E-04	9,30E-07	-211,21	0,00	0,95
125	-2,22E-04	1,29E-06	-172,29	0,00	0,92
126	-1,78E-04	3,23E-06	-55,18	0,00	0,58
127	-1,85E-04	2,37E-06	-78,30	0,00	0,81
129	-2,21E-04	1,81E-06	-122,11	0,00	0,89
131	-2,02E-04	9,52E-07	-211,69	0,00	0,96
132	-1,98E-04	2,05E-06	-96,67	0,00	0,96
162	-2,29E-04	5,87E-06	-39,01	0,00	0,68
163	-2,12E-04	1,46E-06	-145,28	0,00	0,97
164	-2,01E-04	1,44E-06	-139,53	0,00	0,90
165	-1,99E-04	2,29E-06	-86,56	0,00	0,95
167	-2,28E-04	8,29E-07	-275,45	0,00	0,91
168	-2,06E-04	1,63E-06	-126,36	0,00	0,71
181	-2,30E-04	1,62E-06	-141,44	0,00	0,90
183	-1,35E-04	4,36E-06	-30,97	0,00	0,47
187	-2,42E-04	4,08E-06	-59,32	0,00	0,91
194	-2,06E-04	2,91E-06	-70,57	0,00	0,82
195	-1,64E-04	1,03E-06	-159,50	0,00	0,96

TABLE B.2: Temperature coeff. for highest consumption group

Zip	Coeff.	Std. Error	t-value	P value	R square
111	8,16E-05	8,88E-06	9,19	0,00	0,91
112	1,11E-04	2,46E-05	4,51	0,00	0,73
113	1,28E-04	1,93E-05	6,63	0,00	0,83
114	6,87E-04	5,71E-04	1,20	0,23	0,03
115	8,77E-05	6,43E-06	13,63	0,00	0,97
116	1,31E-04	1,66E-05	7,91	0,00	0,95
117	9,38E-05	1,14E-05	8,21	0,00	0,88
118	6,62E-05	1,28E-05	5,19	0,00	0,93
120	6,29E-05	1,06E-05	5,96	0,00	0,92
121	8,10E-05	4,28E-06	18,94	0,00	0,95
122	6,03E-05	2,11E-05	2,86	0,00	0,73
123	6,30E-05	1,09E-05	5,80	0,00	0,95
124	8,29E-05	5,96E-06	13,90	0,00	0,95
125	5,57E-05	8,26E-06	6,74	0,00	0,92
126	1,58E-04	2,07E-05	7,61	0,00	0,58
127	7,92E-05	1,52E-05	5,22	0,00	0,81
129	8,88E-05	1,16E-05	7,66	0,00	0,89
131	7,98E-05	6,11E-06	13,07	0,00	0,96
132	1,00E-04	1,32E-05	7,60	0,00	0,96
162	-1,05E-04	3,76E-05	-2,79	0,01	0,68
163	5,90E-05	9,36E-06	6,30	0,00	0,97
164	8,67E-05	9,21E-06	9,41	0,00	0,90
165	-5,06E-06	1,47E-05	-0,34	0,73	0,95
167	6,25E-05	5,31E-06	11,77	0,00	0,91
168	8,20E-05	1,05E-05	7,83	0,00	0,71
181	7,59E-05	1,04E-05	7,29	0,00	0,90
183	-7,18E-05	2,79E-05	-2,57	0,01	0,47
187	-4,86E-05	2,62E-05	-1,86	0,06	0,91
194	3,56E-05	1,87E-05	1,91	0,06	0,82
195	6,35E-05	6,60E-06	9,62	0,00	0,96

TABLE B.3: Wind speed coeff. for highest consumption group

Zip	Coeff.	Std. Error	t-value	P value	R square
111	0,00392973	1,14E-05	343,8589984	0	0,927184354
112	0,003841326	8,06E-06	476,4068814	0	0,814415538
113	0,003892362	4,47E-06	870,6289537	0	0,946226613
114	0,003937467	6,12E-06	643,7480034	0	0,939724478
115	0,003869728	6,47E-06	598,0102095	0	0,901289331
116	0,003863215	4,43E-06	871,2131042	0	0,949310251
117	0,003845398	5,80E-06	663,0100777	0	0,948282204
118	0,003864402	5,41E-06	713,9829623	0	0,938520575
120	0,003939588	7,55E-06	521,6683692	0	0,930436223
121	0,003897677	3,18E-06	1225,599024	0	0,949350506
122	0,004003719	5,34E-06	749,2816419	0	0,933665035
123	0,003816261	2,81E-05	135,7619054	0	0,659974406
124	0,003866025	8,90E-06	434,4321618	0	0,93511584
125	0,004005311	1,04E-05	384,8158095	0	0,895262813
126	0,003856929	4,92E-06	784,3988751	0	0,934415428
127	0,003856768	2,16E-05	178,8657378	0	0,624703964
128	0,003919143	7,47E-06	524,5021005	0	0,950087568
129	0,003840167	5,06E-06	758,5406289	0	0,939115977
131	0,003865467	1,31E-05	295,2318443	0	0,928915676
132	0,003879182	2,05E-05	188,8343706	0	0,885461953
162	0,004007235	1,04E-05	384,5386121	0	0,936026155
163	0,00384041	8,43E-06	455,6548358	0	0,931976109
164	0,003802851	9,50E-06	400,1086853	0	0,938596091
165	0,00399819	1,45E-05	276,5197558	0	0,931167258
167	0,00404089	5,12E-06	789,1400039	0	0,939021191
168	0,003974257	6,64E-06	598,166921	0	0,870846098
171	0,003934529	2,90E-05	135,4765148	0	0,975264162
181	0,003966451	7,89E-06	502,553388	0	0,884495128
192	0,00405249	7,28E-05	55,63501433	0	0,934082712
193	0,00397236	9,59E-05	41,41315023	0	0,484190538
194	0,003851604	1,95E-05	197,1812479	0	0,842612426
195	0,003904854	1,06E-05	369,0719995	0	0,943934155

TABLE B.4: Intercept for high consumption group

Zip	Coeff.	Std. Error	t-value	P value	R square
111	-2,18E-04	5,14E-07	-423,9125896	0	0,927184354
112	-2,06E-04	3,63E-07	-569,2107883	0	0,814415538
113	-2,13E-04	2,01E-07	-1058,717662	0	0,946226613
114	-2,22E-04	2,75E-07	-805,5323333	0	0,939724478
115	-2,10E-04	2,91E-07	-720,289185	0	0,901289331
116	-2,08E-04	1,99E-07	-1041,389111	0	0,949310251
117	-2,05E-04	2,61E-07	-785,5813357	0	0,948282204
118	-2,12E-04	2,43E-07	-870,0095751	0	0,938520575
120	-2,13E-04	3,40E-07	-626,1690594	0	0,930436223
121	-2,11E-04	1,43E-07	-1473,382067	0	0,949350506
122	-2,20E-04	2,40E-07	-914,0184237	0	0,933665035
123	-1,98E-04	1,26E-06	-156,7818217	0	0,659974406
124	-2,02E-04	4,00E-07	-505,527089	0	0,93511584
125	-2,18E-04	4,68E-07	-465,3356786	0	0,895262813
126	-2,06E-04	2,21E-07	-933,515628	0	0,934415428
127	-1,98E-04	9,70E-07	-203,8690064	0	0,624703964
128	-2,09E-04	3,36E-07	-621,0972867	0	0,950087568
129	-2,06E-04	2,28E-07	-905,8875142	0	0,939115977
131	-2,02E-04	5,89E-07	-343,8102066	0	0,928915676
132	-2,01E-04	9,24E-07	-218,0451676	0	0,885461953
162	-2,18E-04	4,69E-07	-465,9191816	0	0,936026155
163	-1,91E-04	3,79E-07	-502,8530594	0	0,931976109
164	-1,91E-04	4,27E-07	-446,2338427	0	0,938596091
165	-2,18E-04	6,50E-07	-335,5224527	0	0,931167258
167	-2,29E-04	2,30E-07	-996,0199147	0	0,939021191
168	-2,15E-04	2,99E-07	-720,9671923	0	0,870846098
171	-2,20E-04	1,31E-06	-168,5902087	0	0,975264162
181	-2,15E-04	3,55E-07	-604,8374668	0	0,884495128
192	-2,34E-04	3,28E-06	-71,31906812	0	0,934082712
193	-2,10E-04	4,31E-06	-48,73521465	0	0,484190538
194	-1,97E-04	8,78E-07	-224,4202938	0	0,842612426
195	-1,96E-04	4,76E-07	-413,0061144	0	0,943934155

TABLE B.5: Temperature coeff. for high consumption group

Zip	Coeff.	Std. Error	t-value	P value	R square
111	9,11E-05	3,29E-06	27,65705379	0	0,927184354
112	9,45E-05	2,32E-06	40,67687188	0	0,814415538
113	9,22E-05	1,29E-06	71,53542072	0	0,946226613
114	9,67E-05	1,76E-06	54,86448638	0	0,939724478
115	9,25E-05	1,87E-06	49,57379081	0	0,901289331
116	9,01E-05	1,28E-06	70,47401276	0	0,949310251
117	9,00E-05	1,67E-06	53,86461525	0	0,948282204
118	9,89E-05	1,56E-06	63,3643431	0	0,938520575
120	7,63E-05	2,18E-06	35,06747551	0	0,930436223
121	8,57E-05	9,17E-07	93,46659396	0	0,949350506
122	7,11E-05	1,54E-06	46,14958752	0	0,933665035
123	8,46E-05	8,10E-06	10,44122065	0	0,659974406
124	7,75E-05	2,57E-06	30,20933078	0	0,93511584
125	6,65E-05	3,00E-06	22,1615722	0	0,895262813
126	8,94E-05	1,42E-06	63,07759561	0	0,934415428
127	7,01E-05	6,22E-06	11,28296896	0	0,624703964
128	7,44E-05	2,15E-06	34,5297242	0	0,950087568
129	9,45E-05	1,46E-06	64,77026407	0	0,939115977
131	7,79E-05	3,77E-06	20,65463367	0	0,928915676
132	7,09E-05	5,92E-06	11,97766589	0	0,885461953
162	6,70E-05	3,00E-06	22,31033634	0	0,936026155
163	6,01E-05	2,43E-06	24,72141049	0	0,931976109
164	7,24E-05	2,74E-06	26,44353797	0	0,938596091
165	6,97E-05	4,17E-06	16,71714803	0	0,931167258
167	8,04E-05	1,48E-06	54,48634211	0	0,939021191
168	7,13E-05	1,92E-06	37,24857036	0	0,870846098
171	9,47E-05	8,37E-06	11,3162359	0	0,975264162
181	7,19E-05	2,27E-06	31,61659939	0	0,884495128
192	8,66E-05	2,10E-05	4,125389003	4,60E-05	0,934082712
193	6,06E-05	2,76E-05	2,19025008	0,02859705	0,484190538
194	7,11E-05	5,63E-06	12,62220869	0	0,842612426
195	5,24E-05	3,05E-06	17,19532631	0	0,943934155

TABLE B.6: Wind speed coeff. for high consumption group

Zip	Coeff.	Std. Error	t-value	P value	R square
111	0,003935508	7,31E-06	538,579049	0	0,909592263
112	0,003914426	3,44E-06	1137,802519	0	0,914877473
113	0,003944834	2,24E-06	1763,381983	0	0,94497962
114	0,003983911	2,59E-06	1537,662514	0	0,942736456
115	0,003948895	3,90E-06	1012,706048	0	0,939303097
116	0,003889602	3,31E-06	1174,759486	0	0,942896047
117	0,003895336	4,35E-06	894,952482	0	0,940023512
118	0,0039079	3,86E-06	1012,000714	0	0,936841105
120	0,004016354	6,31E-06	636,2697793	0	0,921115276
121	0,003952985	8,24E-06	479,5980817	0	0,848082126
122	0,004152395	6,18E-06	672,0855114	0	0,871349981
123	0,003878082	5,29E-06	733,5364411	0	0,938661225
124	0,003872963	5,94E-06	652,4046068	0	0,926584833
125	0,004052789	8,91E-06	454,9144483	0	0,898939132
126	0,003877363	8,57E-06	452,4513195	0	0,882077494
127	0,003903264	1,10E-05	354,3361855	0	0,784692256
128	0,004025973	7,61E-06	529,3776421	0	0,8971685
129	0,003936683	1,07E-05	367,8163433	0	0,80980484
131	0,003925511	6,25E-06	627,6715853	0	0,952401778
132	0,003871768	1,61E-05	240,6744296	0	0,903265591
162	0,004136223	6,17E-06	670,8676538	0	0,92486641
163	0,003874752	7,73E-06	501,4522808	0	0,93431878
164	0,003821426	1,03E-05	372,288342	0	0,896019192
165	0,004071564	5,89E-06	690,6846861	0	0,922983527
167	0,00418731	8,77E-06	477,1903925	0	0,90569792
168	0,004196596	8,39E-06	499,9891554	0	0,854351788
171	0,003952372	2,32E-05	170,1284651	0	0,967018088
175	0,003936757	2,42E-05	162,9190942	0	0,980403342
181	0,00395126	6,87E-06	575,4933305	0	0,822925984
183	0,003987027	3,66E-05	108,9389295	0	0,814468897
187	0,003868222	2,20E-05	176,2274587	0	0,969459676
193	0,004015686	1,82E-05	221,1608539	0	0,932463727
194	0,004011948	8,03E-06	499,7261572	0	0,911620657
195	0,003976699	9,78E-06	406,7692813	0	0,91661789

TABLE B.7: Intercept for medium consumption group

Zip	Coeff.	Std. Error	t-value	P value	R square
111	-2,21E-04	3,29E-07	-671,9480509	0	0,909592263
112	-2,15E-04	1,55E-07	-1391,83985	0	0,914877473
113	-2,21E-04	1,01E-07	-2193,500327	0	0,94497962
114	-2,29E-04	1,16E-07	-1968,022174	0	0,942736456
115	-2,21E-04	1,75E-07	-1261,224378	0	0,939303097
116	-2,12E-04	1,49E-07	-1425,453413	0	0,942896047
117	-2,12E-04	1,96E-07	-1080,994929	0	0,940023512
118	-2,15E-04	1,74E-07	-1239,116821	0	0,936841105
120	-2,20E-04	2,84E-07	-774,6602448	0	0,921115276
121	-2,16E-04	3,71E-07	-582,6056093	0	0,848082126
122	-2,37E-04	2,78E-07	-851,8211014	0	0,871349981
123	-2,01E-04	2,38E-07	-845,2548762	0	0,938661225
124	-2,03E-04	2,67E-07	-761,6524266	0	0,926584833
125	-2,25E-04	4,01E-07	-561,6770443	0	0,898939132
126	-2,07E-04	3,85E-07	-538,2050568	0	0,882077494
127	-2,07E-04	4,95E-07	-417,2595926	0	0,784692256
128	-2,18E-04	3,42E-07	-638,2239919	0	0,8971685
129	-2,10E-04	4,81E-07	-437,1205247	0	0,80980484
131	-2,07E-04	2,81E-07	-736,9563289	0	0,952401778
132	-2,06E-04	7,23E-07	-284,7517557	0	0,903265591
162	-2,32E-04	2,77E-07	-836,3348072	0	0,92486641
163	-1,95E-04	3,47E-07	-560,3792932	0	0,93431878
164	-1,93E-04	4,62E-07	-417,8902024	0	0,896019192
165	-2,25E-04	2,65E-07	-848,5309162	0	0,922983527
167	-2,45E-04	3,95E-07	-621,1478148	0	0,90569792
168	-2,38E-04	3,77E-07	-631,7592851	0	0,854351788
171	-2,15E-04	1,04E-06	-205,8163716	0	0,967018088
175	-2,06E-04	1,09E-06	-189,9092372	0	0,980403342
181	-2,20E-04	3,09E-07	-712,7143997	0	0,822925984
183	-2,17E-04	1,65E-06	-132,1527537	0	0,814468897
187	-1,83E-04	9,87E-07	-185,3996809	0	0,969459676
193	-2,08E-04	8,16E-07	-254,800509	0	0,932463727
194	-2,17E-04	3,61E-07	-601,7524636	0	0,911620657
195	-2,00E-04	4,40E-07	-454,8244961	0	0,91661789

TABLE B.8: Temperature coeff. for medium consumption group

Zip	Coeff.	Std. Error	t-value	P value	R square
111	9,56E-05	2,11E-06	45,37649353	0	0,909592263
112	9,05E-05	9,92E-07	91,23220327	0	0,914877473
113	9,23E-05	6,45E-07	143,1513409	0	0,94497962
114	9,86E-05	7,47E-07	132,0667007	0	0,942736456
115	9,21E-05	1,12E-06	81,9452537	0	0,939303097
116	9,16E-05	9,54E-07	95,98441866	0	0,942896047
117	8,84E-05	1,25E-06	70,42422863	0	0,940023512
118	9,22E-05	1,11E-06	82,80091563	0	0,936841105
120	6,73E-05	1,82E-06	36,97300871	0	0,921115276
121	7,91E-05	2,38E-06	33,31486928	0	0,848082126
122	6,03E-05	1,78E-06	33,87403934	0	0,871349981
123	7,05E-05	1,52E-06	46,26536793	0	0,938661225
124	7,75E-05	1,71E-06	45,26819595	0	0,926584833
125	6,68E-05	2,57E-06	26,02128716	0	0,898939132
126	8,50E-05	2,47E-06	34,41667888	0	0,882077494
127	7,51E-05	3,18E-06	23,65673316	0	0,784692256
128	6,07E-05	2,19E-06	27,66814471	0	0,8971685
129	7,23E-05	3,08E-06	23,42593262	0	0,80980484
131	6,90E-05	1,80E-06	38,30211075	0	0,952401778
132	8,36E-05	4,64E-06	18,02703976	0	0,903265591
162	5,49E-05	1,78E-06	30,91537683	0	0,92486641
163	5,80E-05	2,23E-06	26,04781991	0	0,93431878
164	7,11E-05	2,96E-06	24,0422408	0	0,896019192
165	6,06E-05	1,70E-06	35,67152103	0	0,922983527
167	6,75E-05	2,53E-06	26,68531399	0	0,90569792
168	4,98E-05	2,42E-06	20,58805083	0	0,854351788
171	7,77E-05	6,70E-06	11,60728403	0	0,967018088
175	6,32E-05	6,97E-06	9,07610686	0	0,980403342
181	8,86E-05	1,98E-06	44,75298535	0	0,822925984
183	7,18E-05	1,05E-05	6,803264346	1,18E-11	0,814468897
187	3,41E-05	6,33E-06	5,394906384	8,42E-08	0,969459676
193	4,15E-05	5,23E-06	7,930528127	2,66E-15	0,932463727
194	6,34E-05	2,31E-06	27,38312309	0	0,911620657
195	3,63E-05	2,82E-06	12,89464101	0	0,91661789

TABLE B.9: Wind speed coeff. for medium consumption group

Zip	Coeff.	Std. Error	t-value	P value	R square
111	0,003891165	1,39E-04	27,90388901	0	0,080158397
112	0,004226764	3,24E-05	130,4045766	0	0,382208399
113	0,004247786	3,19E-05	133,255633	0	0,524137886
114	0,004249293	1,18E-04	35,89696438	0	0,088520218
115	0,004412848	6,85E-05	64,3743683	0	0,218838278
116	0,004059377	2,09E-05	194,4897996	0	0,752887353
117	0,004246423	2,43E-05	174,7918222	0	0,668232014
118	0,004116814	2,43E-05	169,5362045	0	0,63086357
120	0,004446681	2,97E-05	149,6694391	0	0,530905594
121	0,004378589	5,40E-05	81,14700708	0	0,30282407
122	0,004369333	1,80E-05	242,9852616	0	0,663240971
123	0,004029226	6,25E-05	64,46353573	0	0,320048405
124	0,0042648	1,17E-04	36,40830947	0	0,067915233
125	0,004080082	2,35E-05	173,3783456	0	0,585697657
126	0,003985094	6,45E-05	61,81124781	0	0,221502602
127	0,004222267	8,43E-05	50,05945248	0	0,123943253
128	0,004348466	1,39E-05	312,0686166	0	0,758342745
129	0,00413676	1,36E-04	30,4674049	0	0,0528835
131	0,004256378	8,31E-05	51,23561692	0	0,240026855
132	0,004496923	2,04E-04	21,99411872	0	0,085854571
162	0,004485553	2,85E-05	157,2859141	0	0,577848053
163	0,004261316	5,55E-05	76,76797133	0	0,617804437
164	0,004107844	5,78E-05	71,0568219	0	0,591042085
165	0,004196228	1,76E-05	238,1573884	0	0,510218191
167	0,004554179	7,53E-05	60,45543222	0	0,395111528
168	0,004601338	1,65E-05	278,0425554	0	0,683490785
169	0,005594049	1,67E-04	33,59725482	0	0,8720186
181	0,004079864	2,28E-05	179,0334041	0	0,519321
183	0,004574928	1,62E-04	28,19530854	0	0,452645985
187	0,004888741	1,41E-04	34,73544995	0	0,828890026
193	0,0042808	5,39E-05	79,4344618	0	0,578827462
194	0,00437517	4,07E-05	107,4262121	0	0,407505962
195	0,004112609	3,28E-05	125,4777756	0	0,829142485

TABLE B.10: Intercept for low consumption group

Zip	Coeff.	Std. Error	t-value	P value	R square
111	-2,25E-04	6,27E-06	-35,92668794	0	0,080158397
112	-2,30E-04	1,46E-06	-157,6319299	0	0,382208399
113	-2,39E-04	1,43E-06	-167,0314321	0	0,524137886
114	-2,36E-04	5,32E-06	-44,33011591	0	0,088520218
115	-2,30E-04	3,08E-06	-74,65286882	0	0,218838278
116	-2,20E-04	9,38E-07	-234,7890958	0	0,752887353
117	-2,41E-04	1,09E-06	-220,9850579	0	0,668232014
118	-2,26E-04	1,09E-06	-206,5740857	0	0,63086357
120	-2,48E-04	1,34E-06	-185,4770632	0	0,530905594
121	-2,43E-04	2,43E-06	-100,2844324	0	0,30282407
122	-2,56E-04	8,09E-07	-317,0140033	0	0,663240971
123	-2,11E-04	2,81E-06	-74,95915058	0	0,320048405
124	-2,40E-04	5,27E-06	-45,59582404	0	0,067915233
125	-2,31E-04	1,06E-06	-218,121929	0	0,585697657
126	-2,02E-04	2,90E-06	-69,54724937	0	0,221502602
127	-2,30E-04	3,79E-06	-60,6848434	0	0,123943253
128	-2,49E-04	6,27E-07	-397,3187062	0	0,758342745
129	-2,12E-04	6,10E-06	-34,77338604	0	0,0528835
131	-2,43E-04	3,74E-06	-65,01338182	0	0,240026855
132	-2,98E-04	9,19E-06	-32,43207875	0	0,085854571
162	-2,63E-04	1,28E-06	-205,1913877	0	0,577848053
163	-2,34E-04	2,50E-06	-93,65055161	0	0,617804437
164	-2,30E-04	2,60E-06	-88,55134782	0	0,591042085
165	-2,40E-04	7,92E-07	-303,2879691	0	0,510218191
167	-2,71E-04	3,39E-06	-79,87483602	0	0,395111528
168	-2,62E-04	7,44E-07	-352,5067539	0	0,683490785
169	-3,70E-04	7,49E-06	-49,45379737	0	0,8720186
181	-2,34E-04	1,02E-06	-228,0314602	0	0,519321
183	-2,82E-04	7,30E-06	-38,65030492	0	0,452645985
187	-2,64E-04	6,33E-06	-41,69836884	0	0,828890026
193	-2,36E-04	2,42E-06	-97,19078971	0	0,578827462
194	-2,38E-04	1,83E-06	-130,0852423	0	0,407505962
195	-2,23E-04	1,47E-06	-151,0588213	0	0,829142485

TABLE B.11: Temperature coeff. for low consumption group

Zip	Coeff.	Std. Error	t-value	P value	R square
111	1,20E-04	4,02E-05	2,99301687	0,002766892	0,080158397
112	2,07E-05	9,34E-06	2,213796251	0,026848392	0,382208399
113	3,49E-05	9,19E-06	3,795384514	1,48E-04	0,524137886
114	2,69E-05	3,41E-05	0,789722424	0,42969913	0,088520218
115	-3,90E-05	1,98E-05	-1,971675533	0,048660483	0,218838278
116	5,42E-05	6,02E-06	9,005595756	0	0,752887353
117	3,99E-05	7,00E-06	5,69488728	1,25E-08	0,668232014
118	4,71E-05	7,00E-06	6,731520484	1,72E-11	0,63086357
120	-1,15E-05	8,56E-06	-1,340115034	0,180217963	0,530905594
121	1,47E-06	1,56E-05	0,094400358	0,924791948	0,30282407
122	3,29E-05	5,18E-06	6,346348501	2,22E-10	0,663240971
123	4,34E-05	1,80E-05	2,407969458	0,016056582	0,320048405
124	3,14E-05	3,38E-05	0,929819296	0,352472503	0,067915233
125	7,07E-05	6,78E-06	10,41560859	0	0,585697657
126	3,76E-05	1,86E-05	2,025438273	0,042837963	0,221502602
127	2,23E-05	2,43E-05	0,918231425	0,358506218	0,123943253
128	2,33E-05	4,02E-06	5,805846013	6,44E-09	0,758342745
129	1,15E-05	3,91E-05	0,293801047	0,768912743	0,0528835
131	3,95E-05	2,39E-05	1,647625952	0,099452913	0,240026855
132	8,32E-05	5,89E-05	1,411032529	0,158262751	0,085854571
162	1,01E-05	8,22E-06	1,227397214	0,219682724	0,577848053
163	1,86E-05	1,60E-05	1,161697724	0,245409359	0,617804437
164	6,03E-05	1,67E-05	3,62002809	2,97E-04	0,591042085
165	5,43E-05	5,08E-06	10,69646679	0	0,510218191
167	3,89E-06	2,17E-05	0,178988438	0,857950491	0,395111528
168	-2,91E-05	4,77E-06	-6,105920845	1,03E-09	0,683490785
169	-1,14E-04	4,80E-05	-2,381980645	0,017740064	0,8720186
181	7,71E-05	6,57E-06	11,73302478	0	0,519321
183	2,30E-05	4,68E-05	0,491555871	0,623093047	0,452645985
187	-1,19E-04	4,06E-05	-2,934435231	0,00355646	0,828890026
193	1,64E-05	1,55E-05	1,053540699	0,292130276	0,578827462
194	-8,44E-06	1,17E-05	-0,718656738	0,472359267	0,407505962
195	4,20E-05	9,45E-06	4,449758534	8,80E-06	0,829142485

TABLE B.12: Wind speed coeff. for low consumption group

Appendix C

Cluster descriptions

This appendix contains the tables with average values for all clusters per building class. The list below provides extra information about the terms that are used in the tables. All tables are organised in rising order of annual EUI.

Yearly EUI	Annual average EUI [MWh/m^2]
Yealy average	Annual average energy consumption [MWh]
Weekday - weekend	Proportion of consumption during BD/WE [%]
0-5, 6-10,...	Proportion of consumption during indicated hours [%]
Monday, Tuesday,...	Proportion of consumption during indicated day of the week [%]

ClusterNumber	<1925	1926-45	1946-75	1976-05	1976-	2006-	NA	MF Detached	MF in block	MF with other function	SF	Apartment	Size	Yearly EUI	Yearly average	Weekday	Weekend	0-5	6-10	11-14	15-18	19-23	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
6	0,1%	0,0%	0,0%	0,0%	99,9%	0,0%	0,0%		0,1%			99,9%	3	0,024	1,7	92,0	8,0	2,7	20,9	65,7	5,3	5,4	21,8	5,8	5,3	38,1	21,0	4,2	3,8
11	0,0%	0,0%	0,0%	0,0%	0,1%	0,0%	99,9%	0,0%			99,9%	0,1%	3	0,031	211	56,5	43,5	-0,8	-1,0	25,8	49,4	26,5	11,3	9,5	25,7	13,3	-3,3	14,3	29,1
19	0,0%	77,0%	0,5%	20,6%	0,5%	0,0%	1,5%	97,5%	0,5%		1,5%	0,5%	13	0,073	200	65,3	34,7	37,6	23,5	9,9	10,9	18,1	13,8	12,0	12,0	13,2	14,4	16,4	18,3
1	9,6%	9,3%	21,2%	19,6%	0,3%	18,1%	22,0%	53,6%	20,3%	3,8%	22,0%	0,3%	1299	0,075	45	70,2	29,8	25,0	20,9	16,5	16,6	21,1	14,7	13,8	13,5	13,6	14,5	14,6	15,2
17	16,0%	10,0%	22,9%	36,6%	0,1%	14,3%	0,0%	24,3%	53,9%	21,7%	0,0%	0,1%	36	0,089	363	79,1	20,9	16,6	29,2	21,0	17,7	15,5	15,7	14,6	17,1	14,3	17,4	10,5	10,5
8	17,9%	16,6%	22,5%	35,4%	0,0%	6,7%	0,9%	43,2%	52,0%	3,9%	0,9%	0,0%	984	0,095	420	70,5	29,5	22,2	23,4	16,2	17,2	20,9	14,8	14,0	13,7	13,7	14,3	14,5	15,0
14	15,1%	18,8%	23,4%	34,7%	0,1%	7,4%	0,5%	42,5%	47,2%	9,7%	0,5%	0,1%	341	0,108	359	70,5	29,5	18,9	25,2	17,3	18,5	20,2	14,8	14,0	13,7	13,7	14,3	14,5	15,0
18	11,4%	17,8%	39,4%	24,6%	0,6%	6,3%	0,0%	48,6%	44,7%	6,0%	0,0%	0,6%	184	0,122	321	70,6	29,4	18,1	20,7	18,2	20,3	22,8	14,6	14,0	13,7	13,8	14,4	14,6	14,9
0	3,4%	4,6%	64,1%	25,7%	0,0%	1,6%	0,8%	62,5%	33,1%	3,6%	0,8%		457	0,137	1252	70,3	29,7	21,3	22,1	16,8	18,1	21,8	14,6	14,0	13,7	13,7	14,3	14,6	15,1
5	39,0%	23,2%	26,5%	9,1%	0,0%	0,8%	1,4%	34,8%	57,3%	6,6%	1,4%	0,0%	1737	0,137	393	70,5	29,5	24,6	22,4	15,5	16,3	21,2	14,7	14,1	13,7	13,7	14,3	14,5	14,9
2	19,6%	22,3%	20,4%	1,3%	0,1%	0,0%	36,3%	27,7%	31,9%	4,0%	36,3%	0,1%	1326	0,139	50	70,3	29,7	25,0	20,8	16,5	16,7	21,1	14,7	14,0	13,6	13,7	14,3	14,6	15,1
15	6,5%	12,3%	53,7%	20,7%	0,0%	1,1%	5,7%	65,2%	25,4%	3,7%	5,7%		621	0,147	1352	70,4	29,6	23,8	22,3	15,7	16,8	21,3	14,7	14,0	13,7	13,7	14,3	14,6	15,0
13	43,8%	35,5%	16,0%	4,0%	0,0%	0,3%	0,4%	19,9%	76,2%	3,5%	0,4%	0,0%	1495	0,148	326	70,6	29,4	23,1	23,7	16,3	16,4	20,5	14,7	14,1	13,8	13,7	14,3	14,5	14,9
10	3,0%	8,2%	60,8%	24,7%	0,0%	0,0%	3,3%	76,6%	16,5%	3,6%	3,3%		281	0,152	2745	70,4	29,6	22,9	22,1	16,2	17,3	21,6	14,6	14,0	13,7	13,7	14,3	14,6	15,1
7	19,5%	35,0%	40,0%	4,8%	0,0%	0,6%	0,1%	42,5%	54,7%	2,6%	0,1%	0,0%	1372	0,156	391	70,5	29,5	22,3	22,1	16,4	17,6	21,6	14,7	14,1	13,8	13,7	14,3	14,6	14,9
3	0,0%	16,7%	55,0%	26,2%	0,0%	0,0%	2,2%	80,3%	10,2%	7,3%	2,2%		49	0,171	6061	70,5	29,5	23,1	22,0	16,2	17,1	21,6	14,7	14,0	13,7	13,7	14,3	14,5	15,0
9	10,6%	46,0%	37,0%	3,7%	0,0%	0,7%	1,9%	64,6%	28,7%	4,8%	1,9%	0,0%	1052	0,214	280	70,4	29,6	24,4	21,7	15,8	16,7	21,5	14,7	14,0	13,7	13,7	14,3	14,6	15,0
4	0,0%	0,0%	100,0%	0,0%	0,0%	0,0%	0,0%	100,0%					4	0,323	17461	70,3	29,7	23,7	21,2	15,7	17,1	22,2	14,5	14,0	14,0	13,8	14,1	14,6	15,1
12	9,7%	18,4%	46,8%	22,8%	0,0%	2,1%	0,2%	67,7%	29,7%	2,4%	0,2%		61	0,397	895	70,6	29,4	22,8	22,2	16,3	17,2	21,5	14,7	14,1	13,8	13,7	14,3	14,5	14,9
16	0,0%	16,7%	75,9%	7,4%	0,0%	0,0%	0,0%	76,9%		23,1%			8	1,338	2785	70,1	29,9	22,8	21,8	16,2	17,5	21,7	14,7	13,9	13,6	13,7	14,2	14,6	15,2

TABLE C.1: Clusters for residential buildings

ClusterNumber	<1925	1926-45	1946-75	1976-05	2006-	NA	Office and Shop	Office with other function	Warehouse with office	Hotel	Size	Yearly EUI	Yearly average	Weekday	Weekend	0-5	6-10	11-14	15-18	19-23	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
17	4,2%	0,9%	49,8%	33,0%	1,8%	10,3%	77,7%	12,1%	5,2%	5,1%	37	0,072	823	72,3	27,7	14,8	29,3	21,7	19,3	14,8	15,5	14,1	13,5	14,0	15,1	13,9	13,9
5	35,3%	0,6%	42,8%	16,1%	0,0%	5,3%	81,5%	13,1%	5,3%		14	0,077	537	83,3	16,7	11,6	31,7	27,1	18,5	11,1	18,2	16,5	15,6	16,0	17,1	8,4	8,3
0	7,7%	4,6%	39,1%	41,4%	0,0%	7,2%	80,9%	11,9%	7,2%		85	0,084	883	80,4	19,6	18,5	28,3	20,9	17,9	14,4	17,3	16,0	15,3	15,5	16,5	9,6	10,0
12	2,5%	12,9%	47,3%	37,3%	0,0%	0,0%	100,0%				26	0,090	3626	78,4	21,6	21,3	26,5	19,2	16,9	16,1	16,8	15,5	14,9	15,0	16,1	10,8	10,8
1	5,7%	10,9%	28,2%	39,0%	1,4%	14,7%	71,3%	13,9%	14,7%		118	0,092	748	77,5	22,5	21,3	26,3	19,3	16,8	16,3	16,6	15,4	14,7	15,0	15,9	11,1	11,3
19	25,3%	10,7%	20,8%	29,4%	0,0%	13,7%	71,5%	14,8%	12,6%	1,1%	150	0,093	712	73,0	27,0	25,2	22,7	16,5	16,1	19,6	15,5	14,5	14,0	14,1	15,0	13,3	13,7
4	9,2%	5,3%	17,7%	30,6%	0,0%	37,3%	48,3%	14,3%	34,6%	2,8%	40	0,093	321	73,0	27,0	25,4	25,4	16,9	14,9	17,5	15,8	14,4	13,9	13,9	15,0	13,3	13,7
3	21,6%	10,4%	31,3%	29,5%	0,8%	6,3%	67,8%	25,8%	5,9%	0,5%	177	0,096	662	75,1	24,9	22,9	24,5	18,0	16,6	17,9	15,9	15,0	14,3	14,5	15,4	12,3	12,6
9	11,5%	5,2%	28,7%	47,2%	0,0%	7,4%	79,7%	12,9%		7,4%	31	0,100	3143	73,3	26,7	18,9	24,1	19,6	18,9	18,5	15,5	14,4	13,9	14,2	15,3	13,4	13,3
7	18,4%	11,5%	29,4%	30,4%	3,3%	7,0%	65,3%	27,7%	7,0%		93	0,102	598	73,1	26,9	18,8	24,8	19,8	18,6	18,0	15,5	14,4	14,0	14,1	15,1	13,4	13,5
18	18,2%	9,0%	12,7%	23,0%	0,0%	37,1%	44,9%	18,0%	35,5%	1,6%	148	0,108	416	70,3	29,7	27,9	21,3	14,5	15,2	21,1	15,0	14,0	13,4	13,5	14,4	14,6	15,1
2	9,4%	7,0%	46,1%	35,7%	0,0%	1,9%	79,4%	18,6%	0,6%	1,3%	26	0,111	900	73,8	26,2	12,8	21,9	23,2	23,5	18,6	15,2	14,2	14,1	14,6	15,7	13,5	12,7
13	33,0%	14,2%	18,4%	15,6%	1,2%	17,7%	57,7%	24,6%	16,5%	1,2%	230	0,116	355	71,0	29,0	26,1	21,8	15,6	15,8	20,7	15,0	14,1	13,7	13,7	14,5	14,3	14,7
8	19,8%	7,9%	13,3%	19,2%	0,0%	39,8%	43,4%	16,7%	8,6%	31,3%	111	0,119	781	70,7	29,3	22,7	24,2	16,4	16,7	20,1	14,8	14,0	13,6	13,7	14,5	14,6	14,8
11	5,3%	4,6%	6,3%	17,0%	0,0%	66,8%	20,8%	12,4%	2,8%	64,0%	47	0,119	835	70,4	29,6	20,9	27,4	15,3	16,0	20,4	14,3	13,8	13,6	13,8	14,8	14,9	14,7
6	46,9%	10,7%	15,9%	14,0%	3,0%	9,5%	60,9%	29,7%	5,4%	4,1%	188	0,122	429	70,9	29,1	24,4	22,1	16,6	16,5	20,4	15,0	14,0	13,7	13,7	14,5	14,4	14,7
14	18,5%	15,5%	28,2%	22,2%	0,0%	15,5%	67,9%	16,6%	7,8%	7,8%	59	0,125	3233	72,2	27,8	24,5	22,8	16,2	16,3	20,2	15,2	14,3	13,9	14,0	14,9	13,8	14,0
10	35,0%	18,9%	29,7%	11,6%	0,7%	4,2%	70,5%	25,3%	2,2%	2,0%	90	0,131	620	72,2	27,8	21,6	21,7	17,9	18,0	20,9	15,1	14,3	13,9	14,0	14,9	14,0	13,8
16	29,9%	25,8%	13,6%	17,3%	0,0%	13,5%	86,5%			13,5%	7	0,142	10767	72,9	27,1	24,1	23,2	17,0	16,3	19,5	15,4	14,5	14,0	14,1	15,0	13,4	13,7
15	0,0%	0,0%	51,8%	0,0%	0,0%	48,2%		51,8%	48,2%		2	1,446	768	74,6	25,4	24,3	24,1	17,1	16,4	18,2	16,3	14,8	14,1	14,3	15,3	12,5	13,0

TABLE C.2: Clusters for commercial buildings

ClusterNumber	Kyrkor	Samlingslokaler	Skolor	Sportanläggningar	Size	Yearly EUI	Yearly average	Weekday	Weekend	0-5	6-10	11-14	15-18	19-23	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
13		18%	79%	3%	8	0,065	313,57	83,9	16,1	10,9	33,0	27,5	17,4	11,2	17,8	16,8	16,0	16,5	16,9	8,2	7,8
6	2%	35%	23%	40%	23	0,113	714,08	71,7	28,3	17,7	26,1	20,0	18,0	18,2	15,1	14,3	13,7	13,9	14,7	13,9	14,4
0	5%	17%	60%	18%	78	0,114	892,93	71,0	29,0	24,8	22,0	16,3	16,4	20,6	14,9	14,1	13,7	13,8	14,6	14,4	14,6
7	1%		99%		39	0,118	681,61	80,4	19,6	19,2	28,5	21,5	16,2	14,6	17,1	16,0	15,3	15,5	16,5	9,6	10,1
3	0%	12%	75%	12%	55	0,121	1127,96	76,5	23,5	21,1	25,2	19,2	17,1	17,4	16,1	15,2	14,6	14,8	15,8	11,7	11,8
14	0%	1%	98%	1%	89	0,130	974,45	74,8	25,2	23,6	24,2	18,0	15,7	18,5	15,7	14,9	14,3	14,5	15,4	12,4	12,8
12	0%		99%	1%	36	0,141	769,34	77,8	22,2	23,0	26,8	19,1	14,9	16,3	16,5	15,5	14,9	15,0	15,9	10,9	11,4
5	3%	5%	92%	1%	79	0,142	872,54	73,2	26,8	25,0	23,1	16,9	15,6	19,3	15,4	14,6	14,1	14,2	15,0	13,2	13,5
2	6%	43%	16%	35%	44	0,161	1100,92	71,5	28,5	21,2	23,4	17,8	17,8	19,9	14,9	14,3	13,8	13,9	14,6	14,1	14,4
1	14%	23%	43%	21%	98	0,164	439,79	70,8	29,2	26,8	21,5	15,3	15,4	20,9	14,9	14,1	13,7	13,7	14,5	14,4	14,7
10	3%	67%	0%	30%	16	0,177	735,81	73,8	26,2	14,0	22,4	20,4	21,8	21,4	15,3	14,4	14,1	14,4	15,7	13,5	12,7
11				100%	7	0,198	470,92	70,4	29,6	33,0	21,7	12,3	11,9	21,1	15,2	13,9	13,4	13,4	14,5	14,9	14,7
4	10%	19%	12%	60%	26	0,224	624,34	70,5	29,5	23,3	19,9	16,2	18,0	22,6	14,6	14,0	13,5	13,9	14,6	14,7	14,7
9		16%		84%	4	0,258	22679,96	71,9	28,1	25,2	22,0	16,4	16,4	20,1	15,0	14,3	14,0	13,9	14,7	13,9	14,2
8		12%	8%	81%	6	0,939	1654,36	72,5	27,5	24,7	22,7	17,1	16,1	19,4	15,1	14,3	14,0	14,3	14,8	13,6	13,9

TABLE C.3: Clusters for public buildings

ClusterNumber	Daghem och Fritidshem	Servicehus	Sjukhus och Sjukhem	Size	Yearly EUI	Yearly average	Weekday	Weekend	0-5	6-10	11-14	15-18	19-23	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0,59%	16,53%	82,88%	30	0,088	1669,33	73,5	26,5	23,6	23,7	17,3	16,4	18,9	15,4	14,4	14,1	14,4	15,2	13,2	13,3
10			100,00%	3	0,093	24996,31	71,3	28,7	25,6	22,6	16,1	15,6	20,0	15,2	14,1	13,7	13,7	14,6	14,3	14,4
3	0,15%	43,47%	56,38%	66	0,099	1189,57	71,0	29,0	24,9	22,6	16,2	16,2	20,1	14,9	14,1	13,7	13,7	14,5	14,4	14,7
1	1,47%	10,25%	88,28%	32	0,110	830,25	70,9	29,1	27,6	21,7	14,7	14,9	21,1	15,0	14,1	13,5	13,7	14,7	14,4	14,7
6		64,71%	35,29%	6	0,110	468,08	72,8	27,2	18,3	26,7	18,5	18,6	17,9	15,3	14,5	13,9	14,1	15,1	13,5	13,8
11	1,34%	60,28%	38,38%	6	0,110	745,12	71,5	28,5	17,1	27,3	20,7	19,0	15,9	15,1	14,2	13,6	13,8	14,8	14,1	14,4
4	17,99%	45,49%	36,52%	10	0,121	245,41	70,7	29,3	22,0	24,0	15,3	18,4	20,2	14,7	14,0	13,8	13,9	14,4	14,5	14,8
12	55,27%		44,73%	32	0,121	131,72	78,4	21,6	21,1	27,3	21,7	14,9	15,1	16,5	15,7	15,1	15,2	15,9	10,6	11,0
13	0,39%	67,74%	31,87%	31	0,126	1209,04	71,5	28,5	22,1	23,9	17,8	16,9	19,3	15,0	14,3	13,8	13,9	14,6	14,1	14,4
5	96,00%		4,00%	19	0,135	124,04	83,5	16,5	17,0	32,4	22,5	15,9	12,3	17,5	18,2	15,6	15,6	16,5	8,0	8,5
9	50,54%	15,54%	33,92%	20	0,148	158,58	77,7	22,3	21,8	25,4	19,4	16,3	17,2	16,2	15,5	15,0	15,1	15,8	11,1	11,3
7	81,41%		18,59%	32	0,149	127,89	74,1	25,9	24,5	24,6	18,7	14,3	17,9	15,6	14,8	14,4	14,4	15,1	12,8	13,1
2	6,13%	31,55%	62,32%	44	0,164	1548,73	71,5	28,5	26,2	22,2	15,6	15,6	20,4	15,0	14,2	13,7	13,9	14,7	14,1	14,4
14	0,89%	47,95%	51,16%	30	0,205	1297,45	71,1	28,9	24,4	22,6	16,5	16,3	20,2	14,9	14,1	13,7	13,9	14,5	14,3	14,6
8	87,57%	12,43%		6	0,365	182,94	72,8	27,2	24,2	22,5	17,6	15,2	20,5	15,0	14,5	14,1	14,2	15,0	13,6	13,6

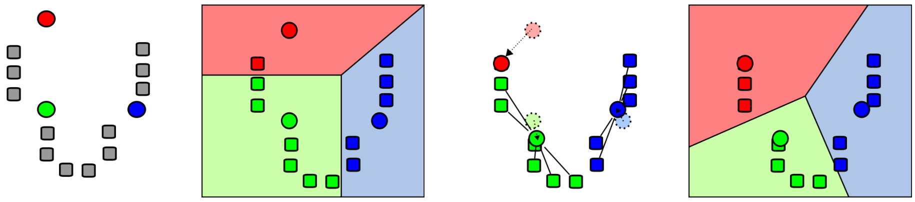
TABLE C.4: Clusters for health & care buildings

ClusterNumber	Fordonsanläggningar	Garage	Verkstäder	Övriga anläggningar	Övriga industriella anläggningar	Grafisk industri	Size	Yearly EUI	Yearly average	Weekday	Weekend	0-5	6-10	11-14	15-18	19-23	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
10			29,3%	33,7%	37,0%		14	0,097	362,13	77,1	22,9	23,0	26,5	18,5	15,8	16,3	16,6	15,3	14,8	14,7	15,6	11,1	11,8
19	11,4%	11,2%	13,3%	38,1%	8,3%	17,7%	25	0,101	756,91	75,6	24,4	22,4	23,9	17,9	17,0	18,8	15,7	14,8	14,8	14,8	15,5	12,2	12,3
11	12,6%		5,9%	36,7%	44,8%		5	0,114	363,45	73,3	26,7	13,9	35,7	22,3	16,4	11,7	15,3	14,3	14,2	14,3	15,2	12,9	13,8
0	6,6%	7,5%	6,9%	71,0%	7,9%		16	0,119	528,23	72,6	27,4	15,7	26,0	21,6	20,3	16,3	15,1	14,4	13,9	14,2	15,1	13,7	13,7
6	52,1%		26,1%	8,3%	13,5%		14	0,119	1000,73	82,4	17,6	16,5	29,8	22,7	18,8	12,2	17,2	16,5	15,7	16,2	16,8	8,9	8,7
13	15,2%	0,2%	31,0%	27,6%	20,9%	5,1%	16	0,125	709,28	78,9	21,1	19,9	27,4	21,8	16,5	14,4	16,9	16,5	14,9	15,2	15,6	10,3	10,7
2	15,9%	41,9%	1,3%	30,9%	10,1%		28	0,135	510,65	70,0	30,0	23,7	23,0	16,0	16,8	20,5	14,9	13,7	13,3	13,5	14,7	14,9	15,1
17			24,2%		75,8%		4	0,144	743,10	79,0	21,0	22,8	18,3	16,9	20,7	21,2	16,5	15,7	15,1	15,5	16,1	9,9	11,1
12	11,5%	4,1%	30,5%	30,2%	23,1%	0,6%	37	0,150	576,37	73,7	26,3	25,5	23,1	16,3	15,8	19,3	15,5	14,6	14,2	14,2	15,1	13,0	13,3
15	9,3%		64,8%	7,5%	18,4%		13	0,155	314,76	73,0	27,0	25,9	25,5	18,0	13,7	16,9	15,5	14,6	14,1	14,0	14,8	13,2	13,8
9	37,9%	0,3%	11,5%	16,3%	30,8%	3,2%	28	0,159	1169,11	72,2	27,8	23,3	22,6	17,4	17,0	19,7	15,1	14,2	13,9	14,0	15,0	13,7	14,0
4	44,3%	5,7%	19,2%	20,5%	10,3%		50	0,160	1216,97	70,9	29,1	25,3	21,2	16,1	16,4	21,0	14,9	14,1	13,7	13,8	14,5	14,3	14,8
18	3,0%	0,5%	6,1%	56,1%	18,6%	15,7%	66	0,160	1519,99	71,2	28,8	27,0	21,7	15,2	15,3	20,8	15,0	14,1	13,7	13,8	14,7	14,2	14,6
14		2,2%	23,4%	53,4%	19,1%	1,9%	41	0,164	603,81	69,6	30,4	28,5	21,0	14,2	14,9	21,5	14,9	13,6	13,2	13,4	14,5	15,0	15,4
16	1,7%		20,6%	46,2%	31,5%		16	0,186	565,29	70,9	29,1	21,0	27,3	17,5	16,5	17,7	14,8	14,0	13,6	13,7	14,7	14,4	14,7
1		25,5%			74,5%		3	0,192	457,42	92,6	7,4	5,3	43,6	34,3	13,5	3,3	17,6	20,9	18,7	18,8	16,6	4,9	2,4
3	1,7%			20,5%	77,8%		7	0,233	390,62	81,3	18,7	10,8	13,6	19,2	33,7	22,7	16,5	16,9	16,6	17,1	14,3	9,7	9,0
8	3,9%	7,5%	6,9%	40,0%	41,7%		25	0,260	287,78	70,8	29,2	19,3	21,5	18,0	18,8	22,3	14,5	13,9	13,7	13,8	14,8	14,6	14,7
5		11,8%	12,0%	70,1%	6,1%		7	1,312	1952,36	66,3	33,7	37,5	17,8	10,5	11,2	23,0	14,0	13,0	12,7	12,7	13,8	16,7	17,0
7				85,0%	15,0%		3	4,131	56552,41	71,0	29,0	25,0	20,9	16,6	16,5	20,9	15,0	14,0	14,0	13,9	14,1	14,4	14,6

TABLE C.5: Clusters for industrial buildings

ClusterNumber	Mätpunkter som bara har varmvatten	Mätpunkter som bara har värme	T-banestationer	Värmeverk, Debiteras ej	Värmeverk	Size	Yearly EUI	Yearly average	Weekday	Weekend	0-5	6-10	11-14	15-18	19-23	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1	100,00%					24	0,037	123,69	70,86	29,14	13,3	28,6	17,1	19,0	21,9	14,6	14,1	14,2	14,1	13,8	14,1	15,0
0	21,34%	12,32%	65,33%	1,02%		32	0,064	97,53	71,52	28,48	22,3	22,9	17,0	17,4	20,3	14,9	14,1	13,9	13,9	14,7	14,1	14,3
7	23,69%	8,97%	62,17%	0,31%	4,86%	36	0,071	120,08	70,75	29,25	26,2	21,4	15,4	16,0	21,1	14,9	14,1	13,6	13,7	14,6	14,5	14,8
4	17,77%		82,23%			8	0,078	83,92	69,03	30,97	21,7	19,2	17,9	19,3	21,9	14,4	13,5	12,8	13,3	15,1	15,5	15,5
2	54,40%		45,60%			11	0,106	131,15	72,08	27,92	12,4	21,9	19,7	22,3	23,8	14,6	14,3	14,1	14,2	14,8	13,9	14,0
9		57,22%		42,78%		2	0,152	3011,57	71,09	28,91	23,9	22,7	16,6	16,4	20,3	14,6	14,2	14,1	13,7	14,4	14,2	14,7
8		21,09%	76,90%	2,01%		15	0,211	413,21	70,86	29,14	25,3	21,4	16,0	16,2	21,1	14,9	14,1	13,6	13,7	14,7	14,5	14,7
5			44,48%	55,52%		2	0,384	1402,45	68,77	31,23	32,2	18,4	13,2	12,8	23,5	14,8	13,7	13,1	13,0	14,2	15,4	15,9
6			100,00%			6	0,509	473,95	70,94	29,06	22,6	22,5	16,6	17,2	21,1	14,8	14,0	13,6	13,7	14,8	14,5	14,5
3			25,98%	74,02%		3	0,774	322,19	70,99	29,01	24,0	22,3	16,1	16,2	21,3	14,5	14,1	13,9	13,8	14,6	14,5	14,5

TABLE C.6: Clusters for other buildings



(A) Initialization (B) Assignment (C) Update (D) Re-assignment

FIGURE 2.1: Illustration of the k-means algorithm with 3 clusters. Source: Weston.pace (2007)

to denote the error between the found (or proposed) model and the data set that is used to construct the model from; variance on the other hand refers to the amount of noise (random variation) that is modeled by the algorithm. The combination of bias and variance gives a measure of how well the model predicts the actual behaviour of the phenomenon (in this case district heating energy consumption) outside the studied data set.

A model with low complexity (i.e. a small number of explanatory variables) does not capture much of the random noise in the data set and thus has a low variance. However, the bias is higher because a too simple model in general does not capture much of the actual behaviour either. Hence, the total model has a high error overall. A model with a too high complexity on the other hand may explain all of the variation that is present in the studied data set. However, this will usually also encompass random variations, which is why the performance on data instances outside the testing set of data might be bad again (high variance, low bias). The tradeoff between bias and variance lies in the minimum in the total error that is encountered between these two extreme cases.

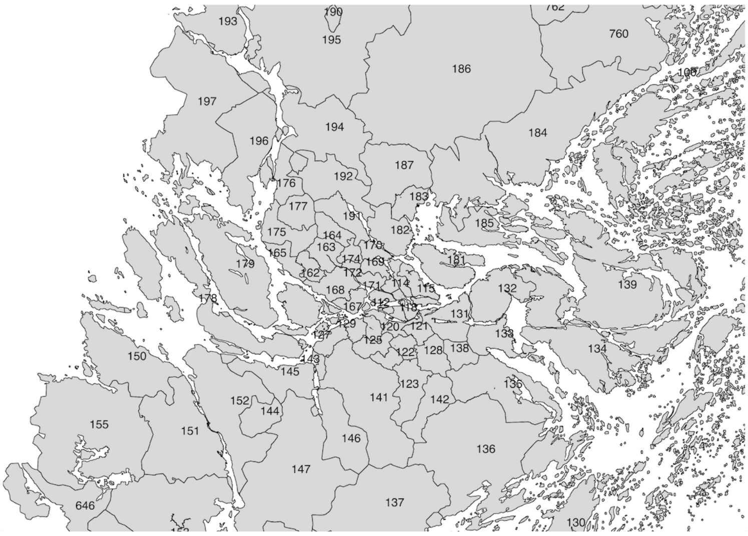


FIGURE 3.2: Map of the areas with the same zip codes, based on the first three digits

- the zip code did exist, but was not included in the map yet.

In case the zip code did not exist, the zip code was replaced by 0 in the KNIME analysis. The same was true for the box zip codes. However, a small number of zip codes with no physical location was used by municipalities surrounding Stockholm (e.g., Lidingö, Upplands Väsby…). In this case, the consumption data under these zip codes were assigned to a real zip code near the center of that community.

In case the zip code did exist according to Posten, but was not included in the map, the base shape file was altered. In some cases, the shape of the missing zip code area could be viewed in Google Maps and hence drawn in QGIS. Sometimes even Google Maps did not show the zip code, and the area had to be approximated using the addresses listed for that zip code by Posten. The drawing of the new areas was performed in QGIS by splitting existing zip code areas to form a new area.
It was observed that most missing zip code areas that were added according to these steps were recently developed or even to-be-developed parts of the city, such as Norra Djurgårdsstaden (Stockholm Royal Seaport) and Hagastaden.

3.4.4 Three digit zip code maps

Because of the irregular size of Stockholm's zip code areas, it will prove useful for the energy maps in this thesis to study the data in larger, more regular areas. Because of the way the zip code system in Sweden works, the areas which share the same three digits of their zip code are suitable for this purpose. An example of these areas is shown in Figure 3.2.

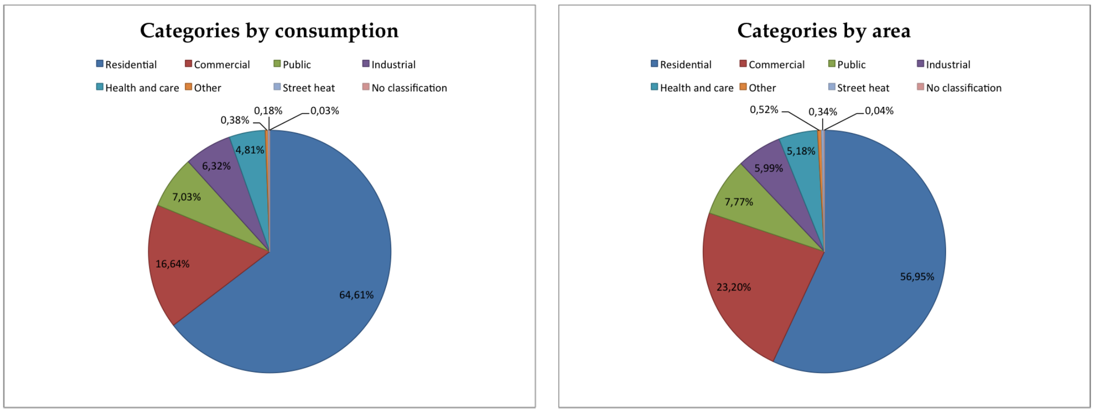
Chapter 4

Results

4.1 Descriptive statistics

This section tries to provide a description of the averages and ranges that are encountered for the different building categories. The results are grouped per building class (see section 3.6.1) and where possible, information about vintage type of the buildings is provided as well.

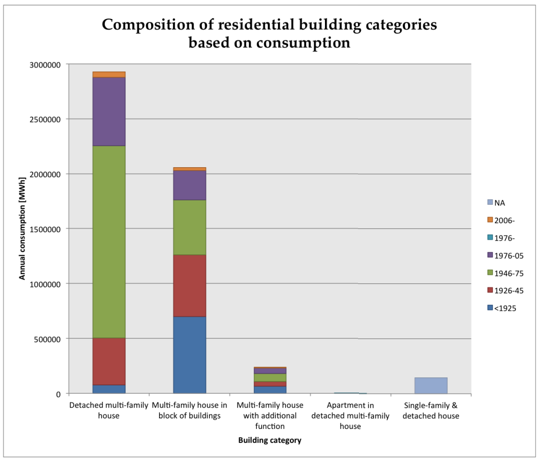
Before putting emphasis on the separate categories in each class, first an overview of all categories is given in the form of pie charts (figure 4.1). Figure 4.1a calculates the proportion of each class based on annual energy consumption, while figure 4.1b uses the building area. Both charts show more or less the same distribution, except for the residential class, which has a higher consumption percentage compared to its area, and the commercial class, where the opposite is true. These differences are present, but less outspoken in the remaining classes and categories.



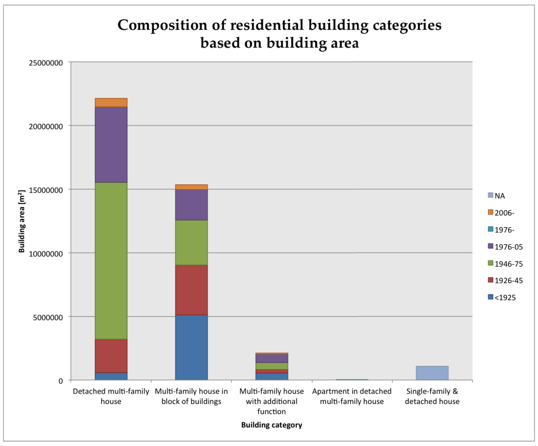
(A) By total annual consumption

(B) By total building area

FIGURE 4.1: Pie charts of building classes



(A) Residential annual consumption



(B) Residential building area

FIGURE 4.4: Composition of residential buildings in terms of categories and vintage

and category, divided by the sum of their areas. This method enlarges the differences between the compositions in figures 4.4a and 4.4b.

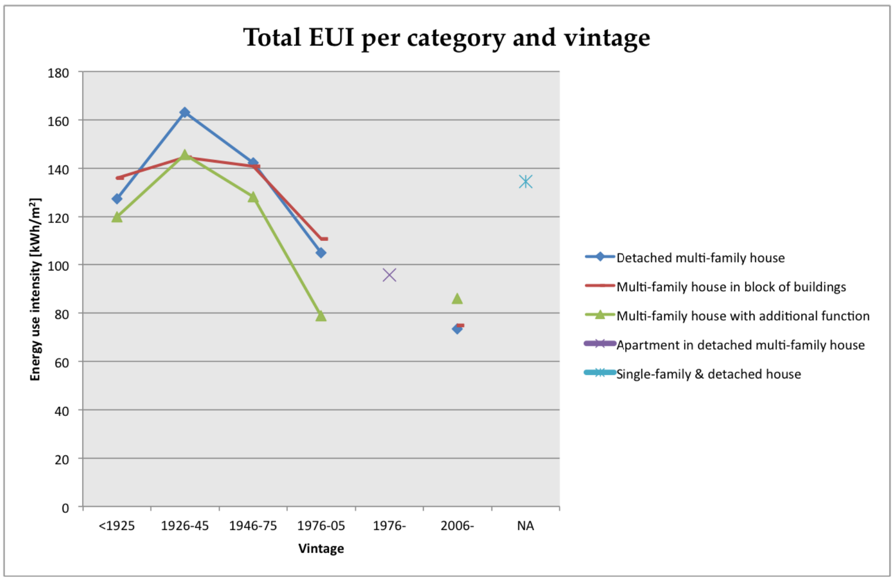


FIGURE 4.5: Total EUI per category and vintage period (expressed in kWh/m^2) for residential buildings

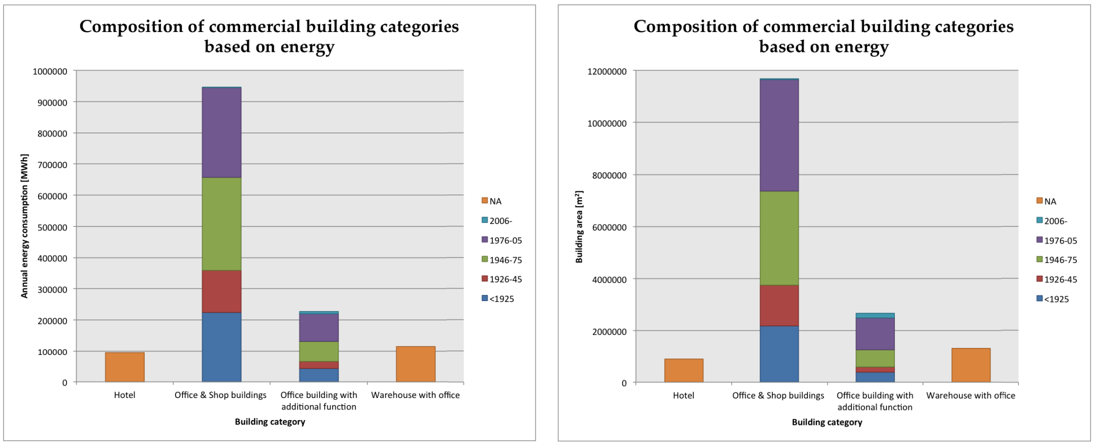
As pointed out by the differing EUI values, the compositions based on energy and area are not directly scalable. An interesting observation is that similar temporary evolutions can be observed for the three largest categories: maximum intensity for the buildings from *1926-45*, then decreasing to lower intensities until the most recent buildings. One exception is the category with additional functions, which remains at the same EUI from 1976 onwards. This category has the lowest EUIs overall, while the buildings in block have a lower peak EUI, but the highest values for *-1925* and *1976-05*.
The apartment category appears separately in the *1976-* vintage period, and performs

according to the EUIs of the other building categories in the same time period. Referring back to the paper by Nässén and Holmberg (2005), this illustrates the stagnation of energy efficiency in the last quarter of the 20th century. The single family houses have a rather high EUI, but it must be remarked that this is an average value for all building periods, which remain unknown.

4.1.2 Commercial buildings

The averages, standard deviation and extreme values for the different categories of commercial values are gathered in table A.2. It is immediately clear that the number of buildings of this type is smaller than the amount of residential buildings. Again, the high variance in total consumption as opposed to a low variance in EUI is apparent. Remark the lower mean EUIs when compared to the residential use intensities. At the same time, the commercial buildings spread a larger floor area than their residential counterparts.

Using the same method as for the residential buildings, the composition of the commercial building class is shown based on energy in figure 4.6a and based on building area in 4.6b. Quite similar compositions are shown again, with a high amount of office and shop buildings, and a smaller amount of these with an additional function. The hotel and warehouse categories are less represented, and have no vintage information moreover. The composition of the two largest categories seem similar in proportion, but the additional function category has more buildings from 2006 onwards pro rata.



(A) Commercial annual consumption (B) Commercial building area

FIGURE 4.6: Composition of commercial buildings in terms of categories and vintage

With the previously described technique, the overall energy use intensities per vintage and category can be calculated. The resultant numbers are showed as a graph in figure 4.7. The two office buildings categories display a very interesting evolution: the oldest buildings start out at an overall EUI around 100-110 kWh/m^2, but quickly decrease to only 40 kWh/m^2 in the newest buildings. These values are all lower than the EUIs for residential buildings. In this case, there is no direct evidence of the efficiency stagnation that Nässén and Holmberg (2005) point out.

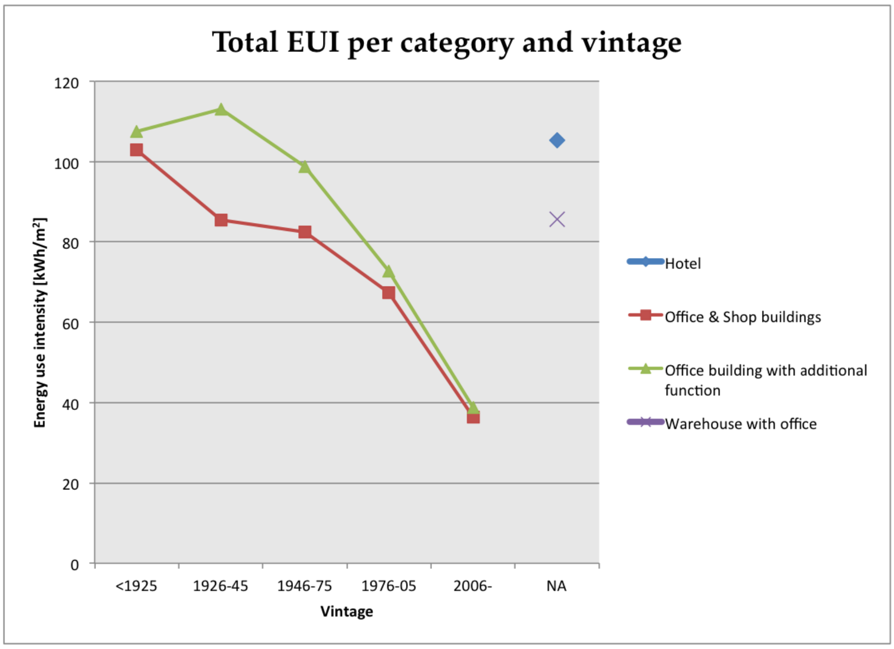


FIGURE 4.7: Total EUI per category and vintage period (expressed in kWh/m^2) for commercial buildings

4.1.3 Health & care buildings

Table A.3 summarizes the descriptive statistics of buildings for the health and care class. The energy use intensities for these categories are situated in the same range as the residential meters. The high annual consumption for hospitals and rather small consumption for daycare buildings follow from the annual energy consumption means.

Furthermore, the respective shares of building categories for this class are presented in figure 4.8.

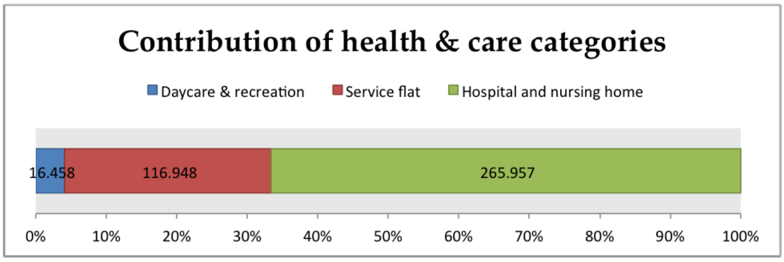


FIGURE 4.8: Shares of the various health care building categories by annual energy consumption [MWh]

4.1.4 Public buildings

The descriptive statistics for the buildings in the public class can be found in table A.4. The higher EUI means are remarkable; furthermore, the spread on the consumption intensities seems to be higher than in the previously studied classes, pointing to a higher heterogeneity within the public building class.

Again, the shares of the constituting categories are presented in figure 4.9

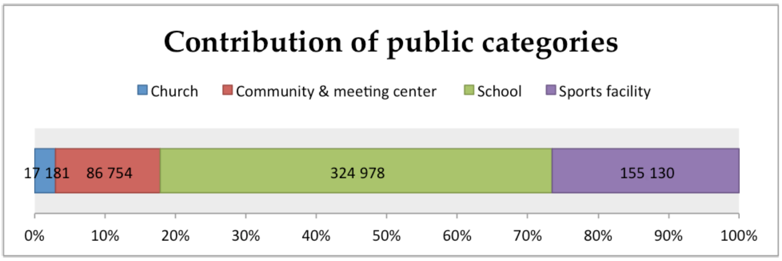


FIGURE 4.9: Shares of public building categories by annual energy consumption [MWh]

4.1.5 Industrial buildings

Table A.5 shows the descriptive statistics for industrial buildings. The heterogeneity of the industrial buildings is reflected in the wide range of mean EUIs per category, Moreover, the standard deviation on the EUI is much higher when compared to the previous results. The respective shares in energy consumption are visualised in figure 4.10.

Remark that there is still one building with an area of 1 m^2. This meter ID has slipped through the outlier removal mechanism, most probably because of a still acceptably low EUI value. Instead the category next to that, with other industrial facilities, appears to have the highest maximum EUI with a surprising value of 12 000 kWh/m^2.

A second comment on the industrial buildings is that the high variation in both EUI and total consumption may be explained by the fact that heat from the DH network is possibly used for other purposes than heating alone. In industrial buildings, steam from the heating system can be used as an input for production processes.

4.1.6 Other buildings

Finally, table A.6 summarizes the descriptive statistics for the remaining building categories. The few buildings with either only warm water, or only heating have rather low (maximum) EUI values. The T-bana (subway) stations have normal consumption figures, compared to the previous cases, although the standard deviation is rather high.

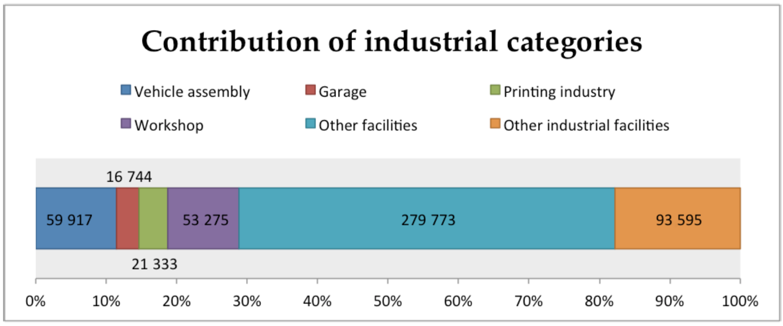


FIGURE 4.10: Shares of industrial building categories by annual energy consumption [MWh]

The 8 district heating plants that have not been identified as outliers display on the one hand rather low EUIs (charged), on the other hand quite high values (not charged).

Street heat is a special class of heating, which shows a remarkably low EUI. The maximal value as well as the standard deviation fall in an equally low category. The buildings with no classification are only incorporated for completeness and will not be analysed further.
The shares in annual energy consumption for the categories within the "other" class, except street heat and non-classified buildings, are shown in figure 4.11

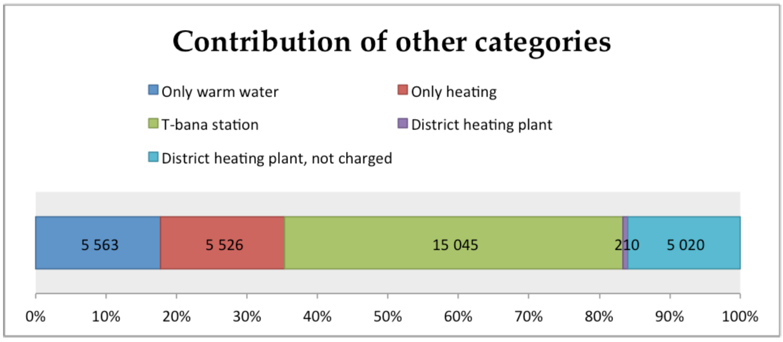


FIGURE 4.11: Shares of other building categories by annual energy consumption [MWh]

4.1.7 *Key findings*

- Residential buildings account for the largest consumption share, followed by commercial buildings
- Buildings from before 1975 are predominant, with 1946-1975 as the largest age group for residential buildings and 1976-2005 for commercial buildings.
- There is a downward time evolution for the EUI in those two classes, but there is a stagnation for the most recent time period
- Very high EUI values occur in the industrial buildings class

4.2 Time-dependence of energy consumption

Improving energy efficiency and reducing GHG emission is not just a question of reducing energy consumption in absolute numbers; one has to keep in mind that by decreasing the peak consumption, the use of polluting peak capacity generators can be made unnecessary. Therefore, it is important to study the instantaneous power demand for district heating as well.
In this case, the power is not known, but the hourly consumption is. The power can thus be approximated by dividing the hourly energy consumption by the time interval:

$$P = \frac{dE}{dt} \approx \frac{\Delta E}{\Delta t}. \tag{4.1}$$

In other words, the hourly consumption in MWh equals the average power demand in MW for that time interval.

4.2.1 Annual

First, a general image of the daily energy consumption throughout 2012 is formed. Figure 4.12 shows the sum of all buildings' consumption, grouped per day. Apparently, consumption is highest around the 30th day of the year, i.e. end January/begin February. As expected, consumption is lowest during summer, but there is a remarkable peak between day 150 and 160.

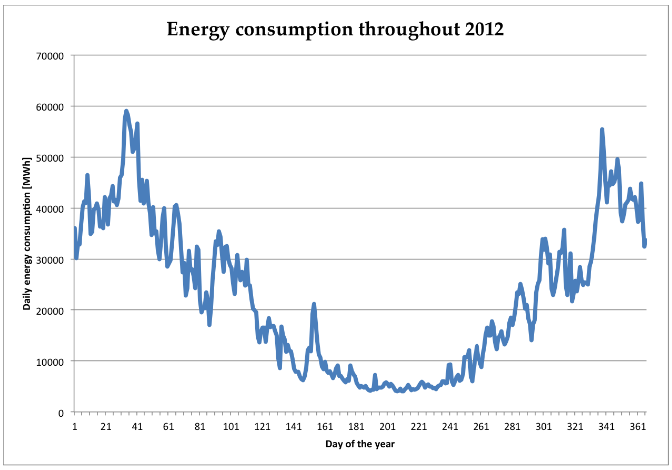


FIGURE 4.12: Daily energy consumption for 2012

A very interesting graph is produced by adding the daily average temperature, but on a negative scale (see figure 4.13). By choosing the scale accordingly, the majority of the observations can be made to coincide, pointing to the fact that the outside temperature has a large influence on the total energy consumption.

This coincidence is less during summer, where the average temperatures increase above about 15 °C. On these days, the consumption does not decrease any longer. The reason for this is that a major part of the buildings does not need heating at these temperatures, but the consumption of hot water continues as usual.

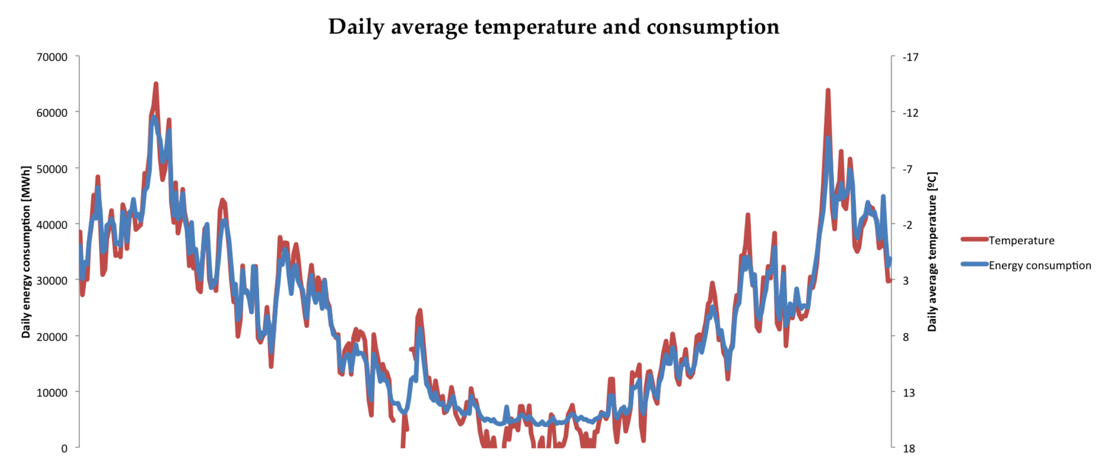


FIGURE 4.13: Daily energy consumption and average temperature

It is however also interesting to see how the consumption varies in the course of one day. Figure 4.14 shows a 3D-plot with the days of the year on one axis, and the hours per day on the other. Cross-sections of the graph along the hour-axis thus show the hourly consumption for one day.

From this figure, it can be seen that most of the days have a peak in consumption during the morning (around 8 AM), while the consumption is on its lowest during nighttime. A remarkable peak during three hours around the 190^{th} day can be observed. Closer investigation has shown that there is one residential building with a consumption of more than 800 MWh per hour at this time, which coincides with the peak. Since it is the only building with such high consumption around that moment it is assumed that these measurements are flawed.

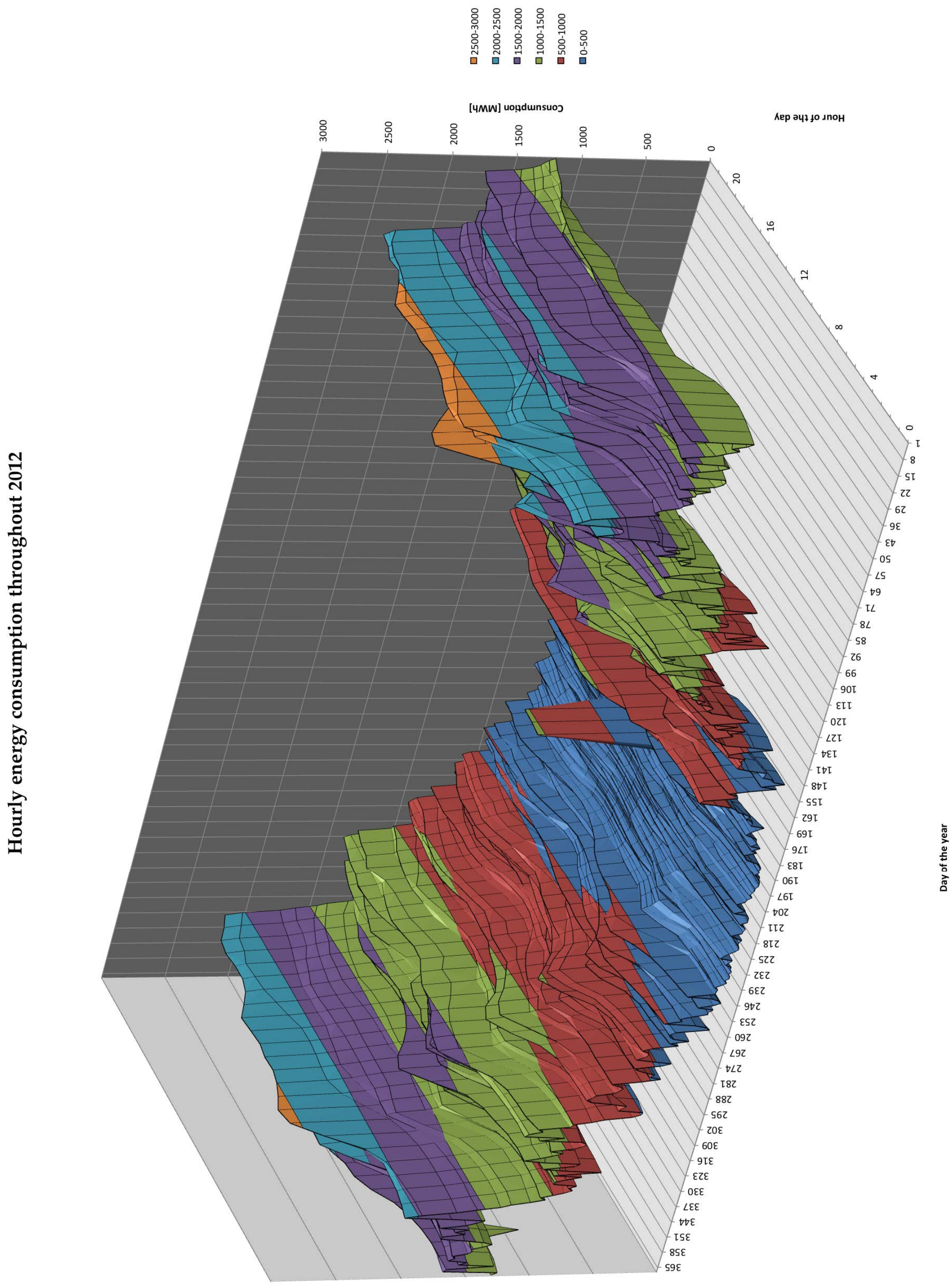


FIGURE 4.14: Hourly consumption for every day in 2012
Note the consumption peaks at 8AM and 6-7PM. The remarkable peak around day 190 at noon can be explained by bad data, since there is one building which consumes the largest part of this peak. Along the day axis, the same evolution as in figure 4.12 can be observed.

4.2.2 Comparison of coldest and warmest day

The observations of the hourly energy consumption are now refined by looking more closely at two individual days. More precisely, the extremes are studied. Since the energy consumption is closely related to outside temperature, this variable is used as a criterion: the coldest and warmest day in the year are considered.

The 4th of February was the day with the lowest average temperature in Stockholm in 2012. Although this was not the day with the highest consumption (3rd of February), their difference in consumption is only minor. Figure 4.15 shows a stacked plot for the hourly energy consumption per class on that day. The largest portion is taken up by residential buildings, followed by commercial buildings. Street heat, other buildings and buildings without classification consume only a marginal amount of energy. The daily peaks at 10 AM and 5-6 PM appear clearly. The hourly average power (or equivalently, aggregate hourly consumption) has a maximum at 2.7 GWh/h.

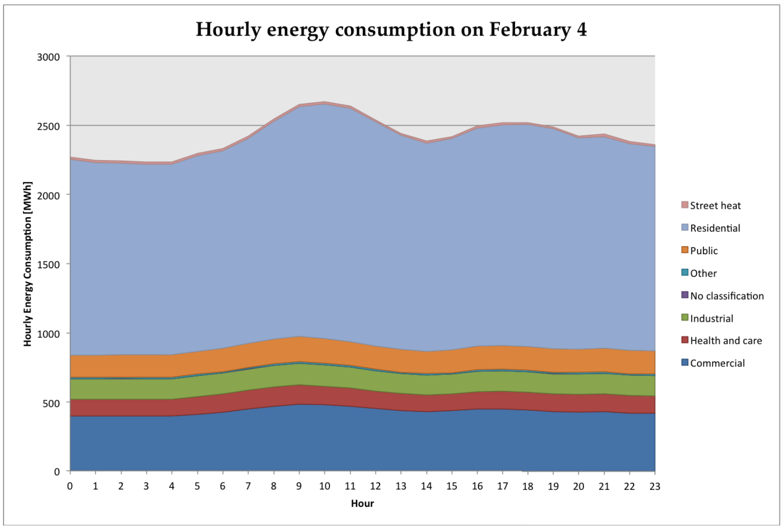


FIGURE 4.15: Energy consumption on the coldest day of 2012

On the warmest day of the year, the 24^{th} of July, the hourly consumption is considerably lower with a peak of only slightly more than 200 MWh at 10 AM. For the rest, the distribution of consumption over the different building classes is similar, with the exception of street heat and no classification, which do not appear in figure 4.16.

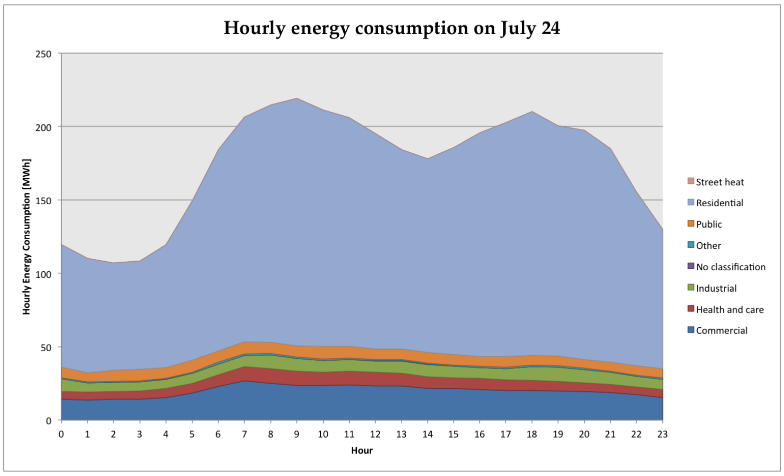
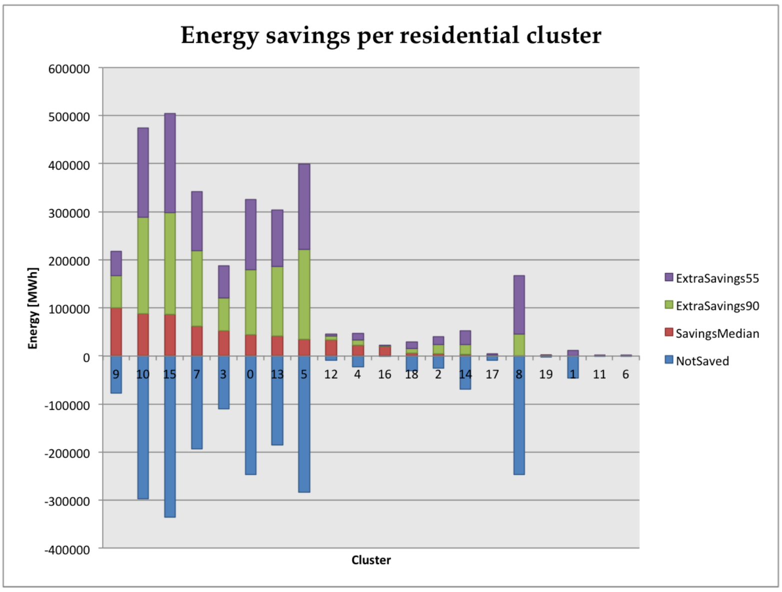


FIGURE 4.16: Energy consumption on the warmest day of 2012

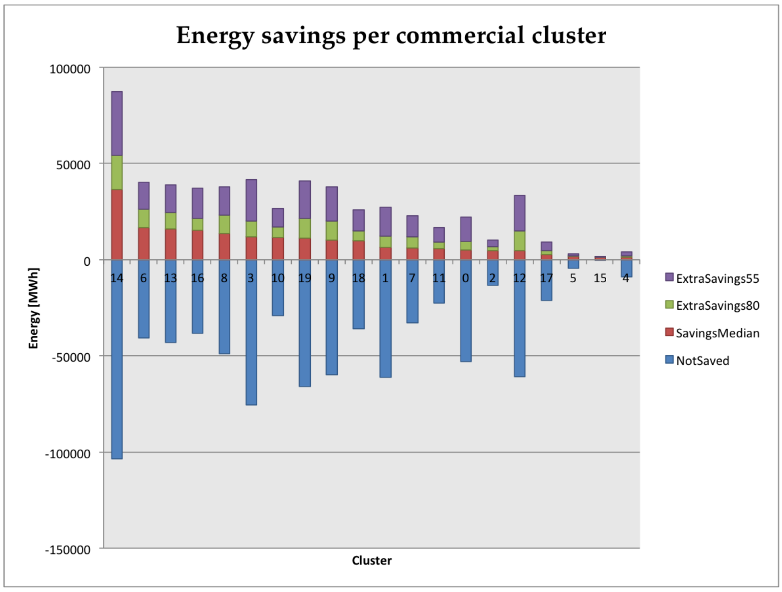
4.3 Energy maps of Stockholm

Figure 4.17 summarises the annual energy consumption data for each 3-digit zip code area. Although the same set of colours is used for all of these maps, one must pay attention to the differing class boundaries. The total annual consumption per class is indicated to provide easier comparison. As expected, residential and commercial buildings have the largest consumption per zipcode area, mostly in and around the centre of Stockholm. For example, the two lowest consumption classes for residential are equal to or greater than the maximal class consumption for most other building classes.

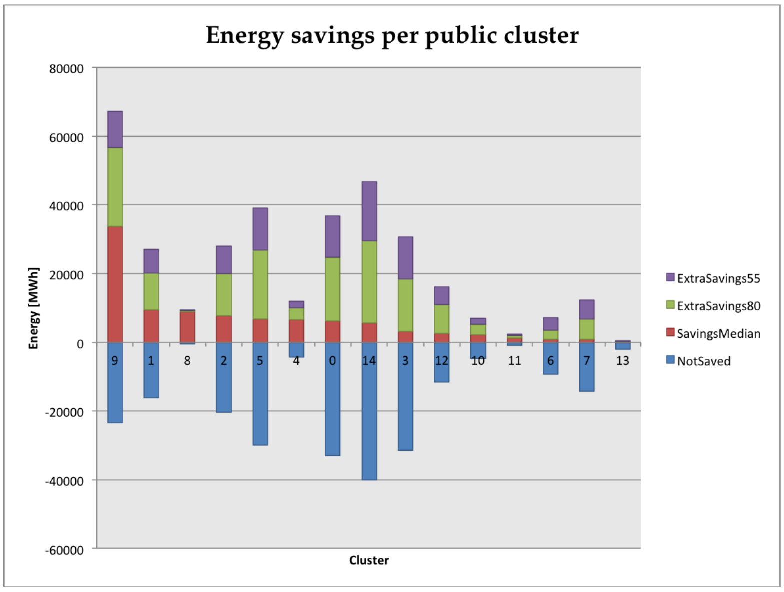
In the first four maps, most of the energy consumption is concentrated in the central areas of Stockholm. The contrast between the centre and the peripheric zip code areas is even greater because of the difference in area between them. It is only for the industrial and other buildings that the consumption in more distant areas is larger than in central areas.



(A)

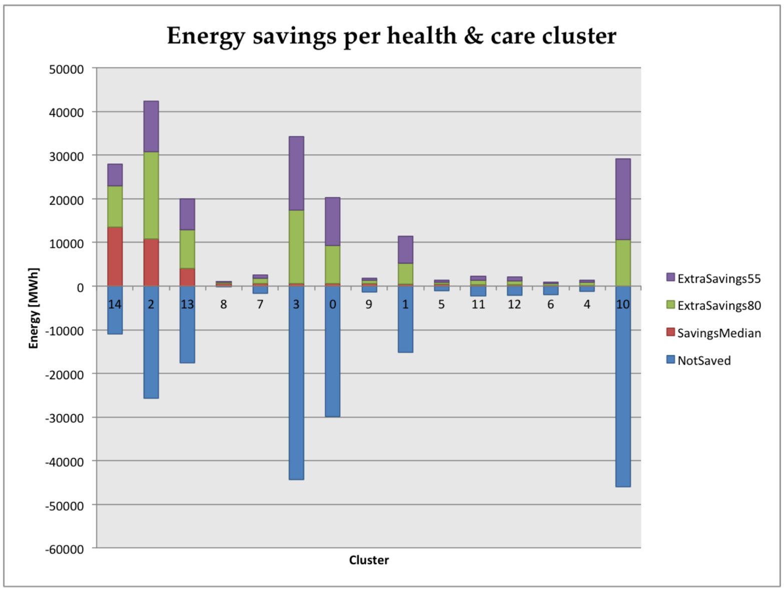


(B)

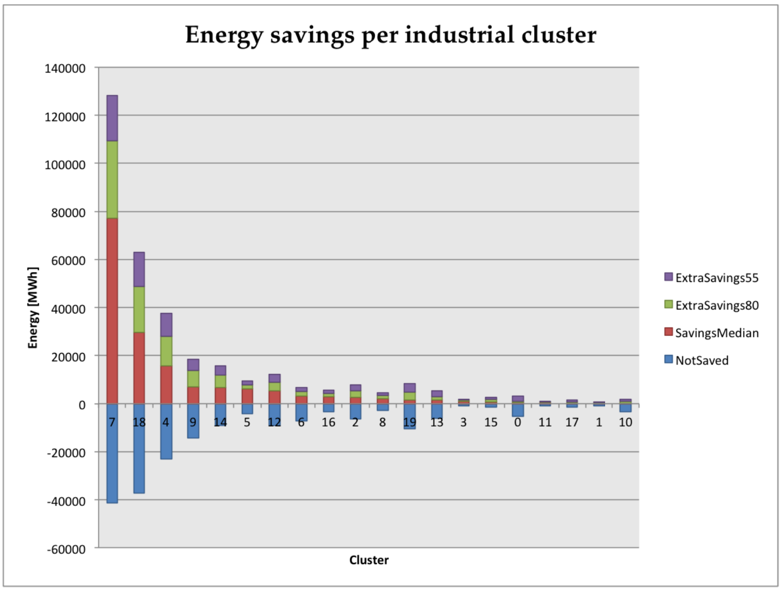


(C)

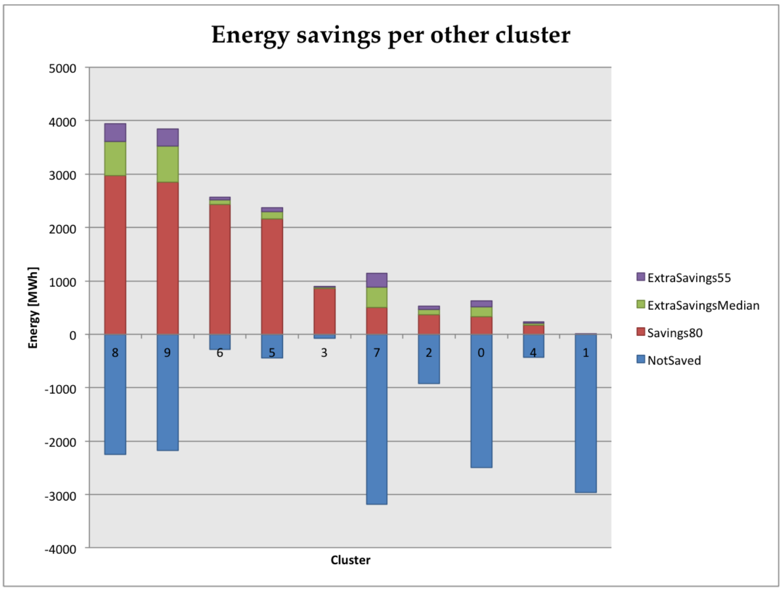
FIGURE 4.19: Cluster prioritisation of energy savings, ordered by low savings



(D)



(E)



(F)

FIGURE 4.19: Cluster prioritisation of energy savings, ordered by low savings *(continued)*

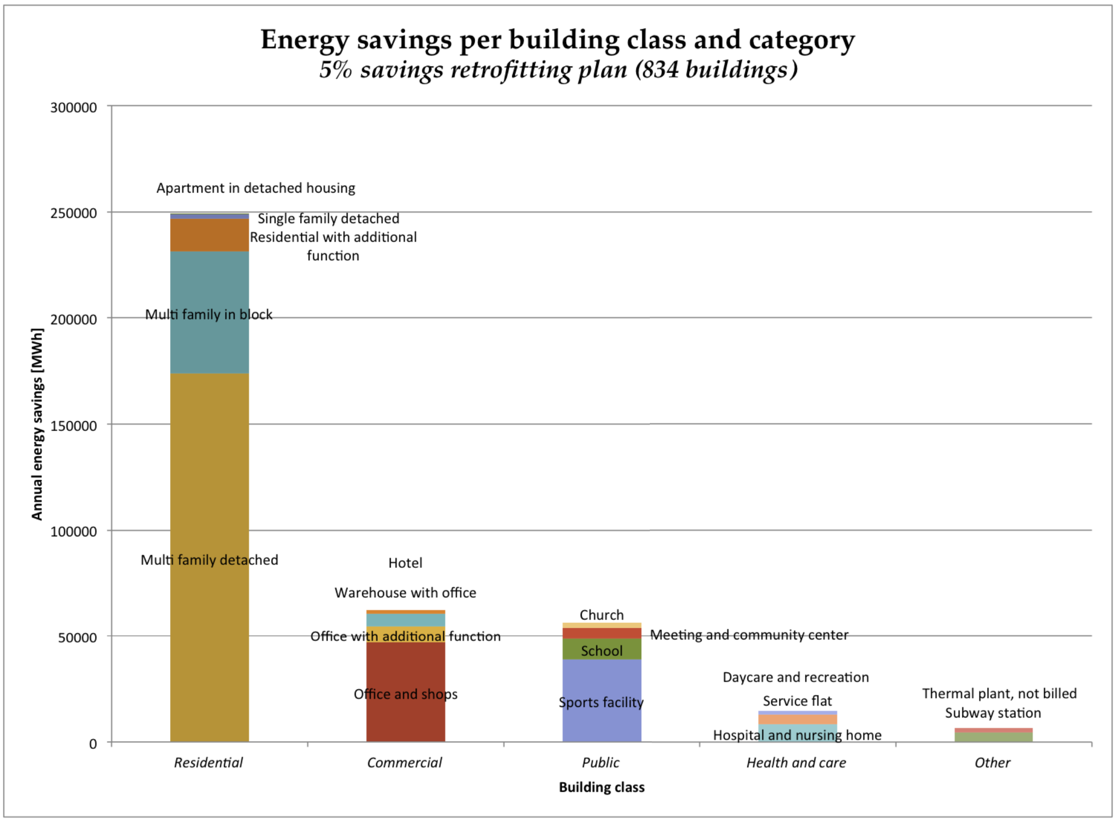


FIGURE 4.27: Energy savings per building category in the 5% energy savings retrofitting plan

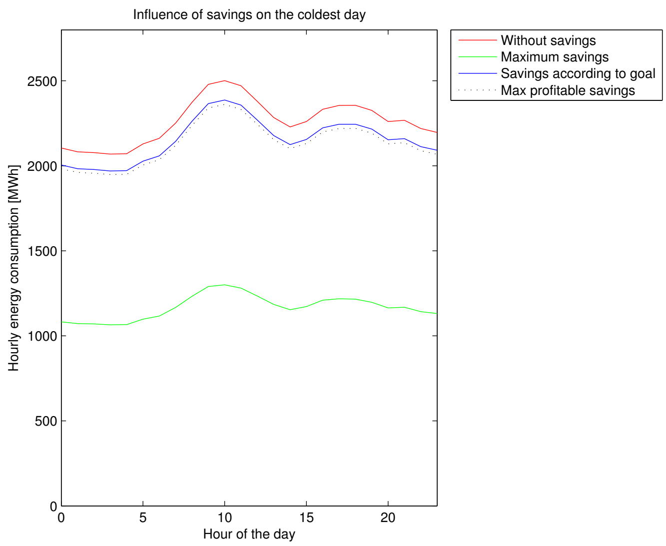


FIGURE 4.28: Heating demand reduction for the coldest day of the year

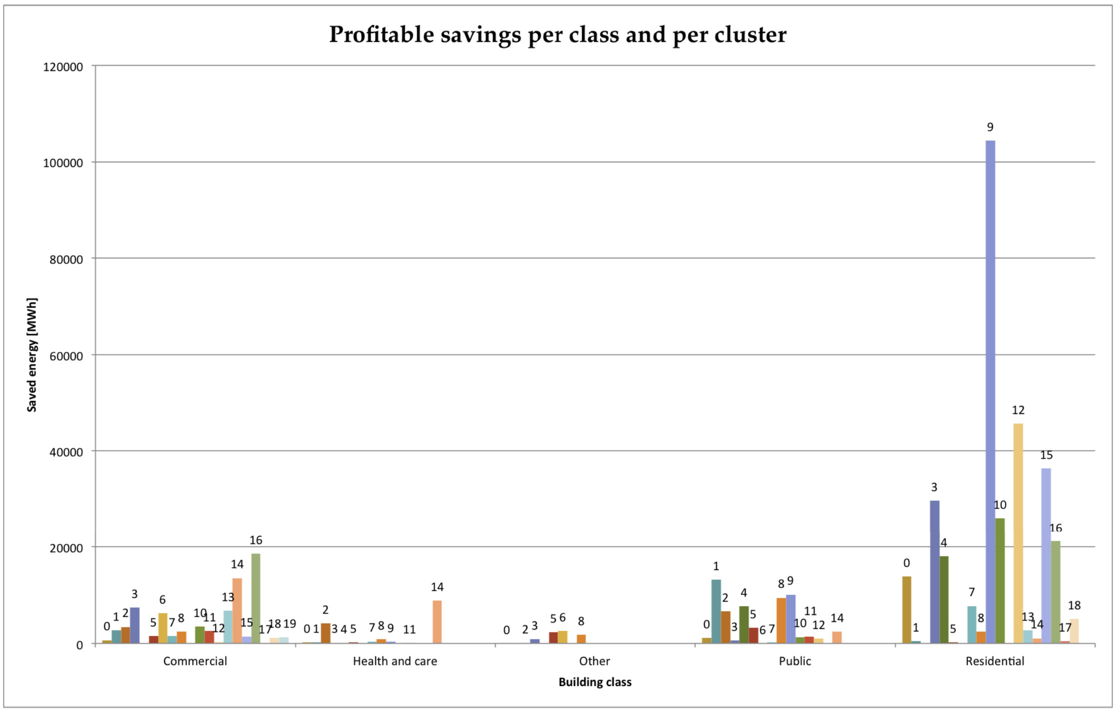


FIGURE 4.29: Maximum profitable savings per building class and cluster

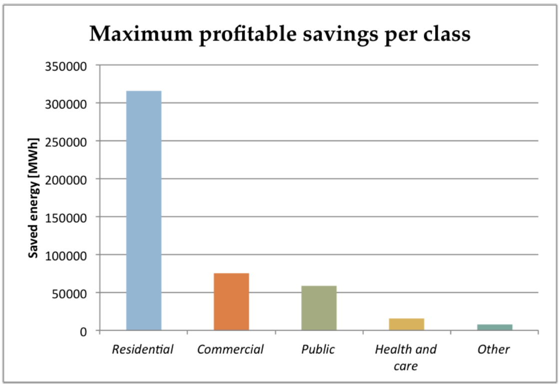


FIGURE 4.30: Maximum profitable savings per class

To finish the results about the maximum profitable savings, figure 4.31 shows a map with the geographical spread of the profitable savings. The largest savings are concentrated in a small number of areas around the city center (Norrmalm & Gamla Stan, Västberga, Södermalm, Bromma and Johanneshov), and one area in the north (Upplands Väsby).

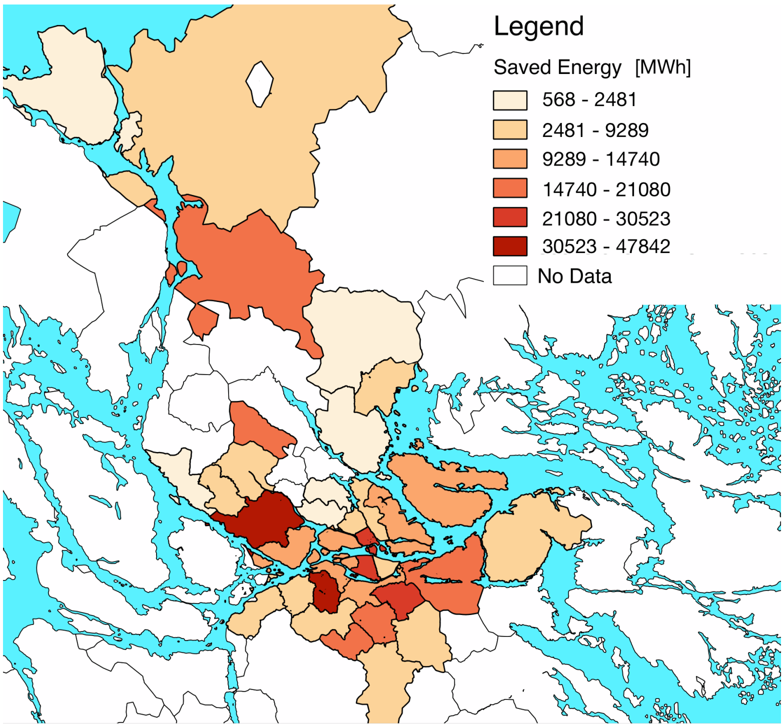


FIGURE 4.31: Map of maximum profitable savings (472 GWh)

4.6.3 Maximum possible savings

The last case that is studied is that in which all buildings that have an EUI greater than 60 kWh/m^2 are retrofitted according to the different types of retrofit measures that are mentioned in the methodology section, irrespective of the economic situation of that investment. In this case, there is an annual savings potential of 3 831 GWh or 49.35% of the total annual consumption (again without industrial buildings). The combined investment that would be needed amounts to 76.9 billion kr, while the combined return because of the saved energy for 15 years would only sum up to 51.7 billion kr.

In this non-economic scenario, 7887 buildings would need a shell renovation, 3396 would be retrofitted with a HRV system, and 576 would receive new thermostats. There are 984 buildings that are below the EUI limit for the retrofitting scenarios. Based on building floor area, 56% of these buildings without savings are commercial, 24% residential, 15% public and the remaining 5% are equally shared by health & care and other buildings.

Again, on figure 4.28 the possible peak demand reduction can be seen; the maximum reduction amounts to 1.2 GWh/h at 11 AM; this reduction is close to 50% of the peak demand.

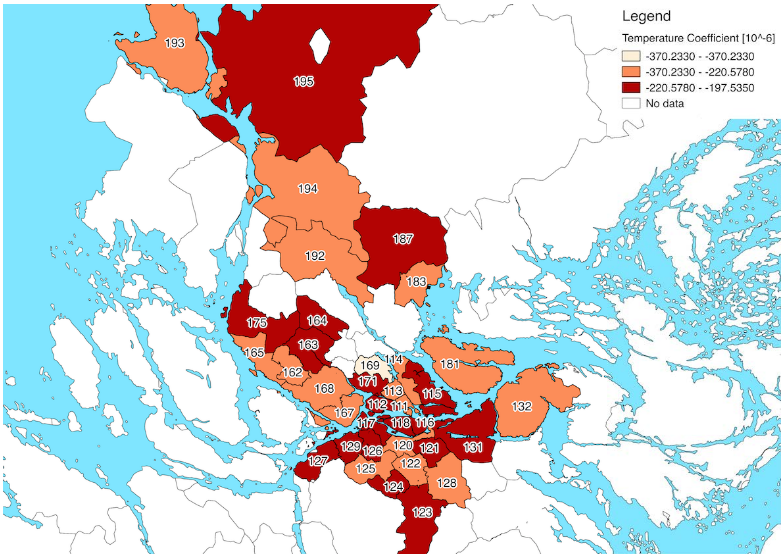


FIGURE 4.37: Choropleth map of the temperature coefficient for spatial regression

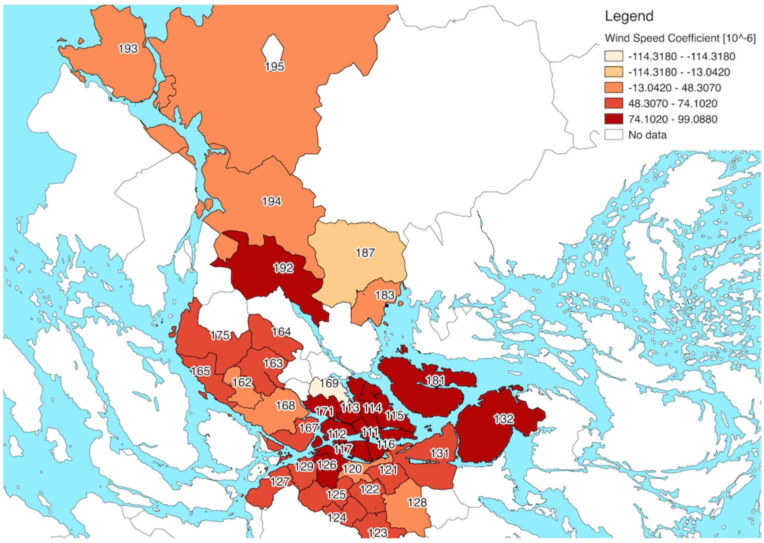


FIGURE 4.38: Choropleth map of the wind speed coefficient for spatial regression

is implemented by the following formula (Navidi, 2008):

$$\bar{R}^2 = R^2 - \left(\frac{k}{n-k-1}\right) R^2, \tag{4.2}$$

with $\bar{R}^2$ the adjusted R^2, k the number of regressors and n the sample size of the regression input. In most cases that are encountered in these regression analyses, n in the denominator is of such a magnitude that the second term in this equation can be safely neglected, and thus the unadjusted R^2 is used.

Apparently, the highest R squared value among the zip codes occurs in 175, where its value is 0.98, and the lowest value is found in 132 with a value of 0.13. The use of this parameter is not clear from the table, therefore it is visualised in another choropleth (figure 4.39).

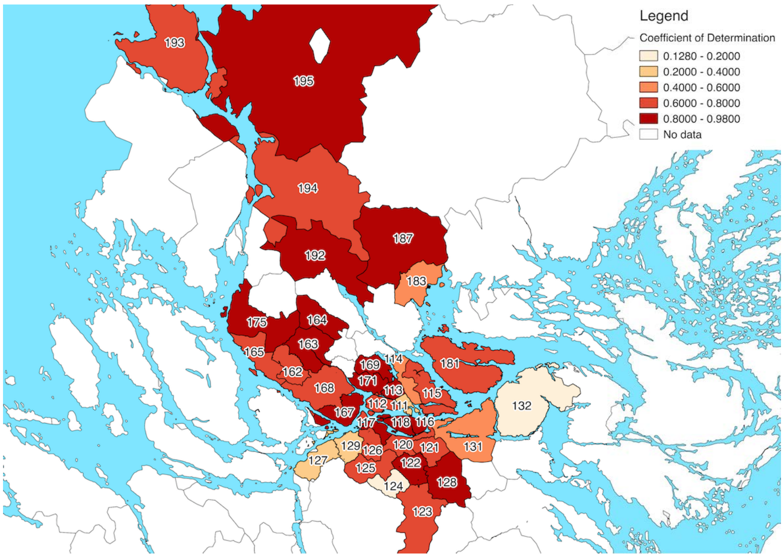


FIGURE 4.39: Choropleth map of the coefficient of multiple determination (R squared) for spatial regression

4.7.2 Spatial regression grouped by consumption

In order to check for different weather dependence for residential buildings with different consumption intensities, the building stock was divided over four groups. As described in paragraph 3.8.2, these four groups can be characterised as buildings with highest, high, medium and low consumption. In order not to overload the report too much with tables, the comprehensive regression results are included in the tables in appendix B. Only the most important results are treated in the current section.

A quick look through the resulting tables in appendix B already tells a lot about the quality of the regression. For most of the zip codes and coefficients, the P values are equally low as in the previous section (none above the 0.05 required for the significance level of 5%), while the R squared values are in general high to very high. Some attention must be paid here: since the regression becomes more and more split up, the number of observations becomes lower, which could imply that the assumption that $R^2 \approx \bar{R}^2$ might not be as accurate as before.
However, for the coefficients in the low consumption group, the R squared values are generally lower. While the P values for the intercept and temperature relation are still 0, now more than half of the zip codes have a wind speed coefficient with a P value that approaches or surpasses 0.05. This indicates that the results for the low consumption class are less accurate and should not be used to draw conclusions.

In order to visualise the difference in weather dependence for the four classes, they have been plotted in the figures below. Figure 4.40 shows the change of the EUI* intercept with respect to the zip code, and grouped per consumption class. Strikingly, the intercept of the low consumption group is in general the highest. However, as remarked in the previous paragraph, the low consumption group displays rather low coefficients of determination. The intercepts for the medium and high consumers seem to be more or less equal, but slightly higher for the medium group. The intercept for the highest consumption group is highly variable, but is often found near the intercept of the medium and high consumers.

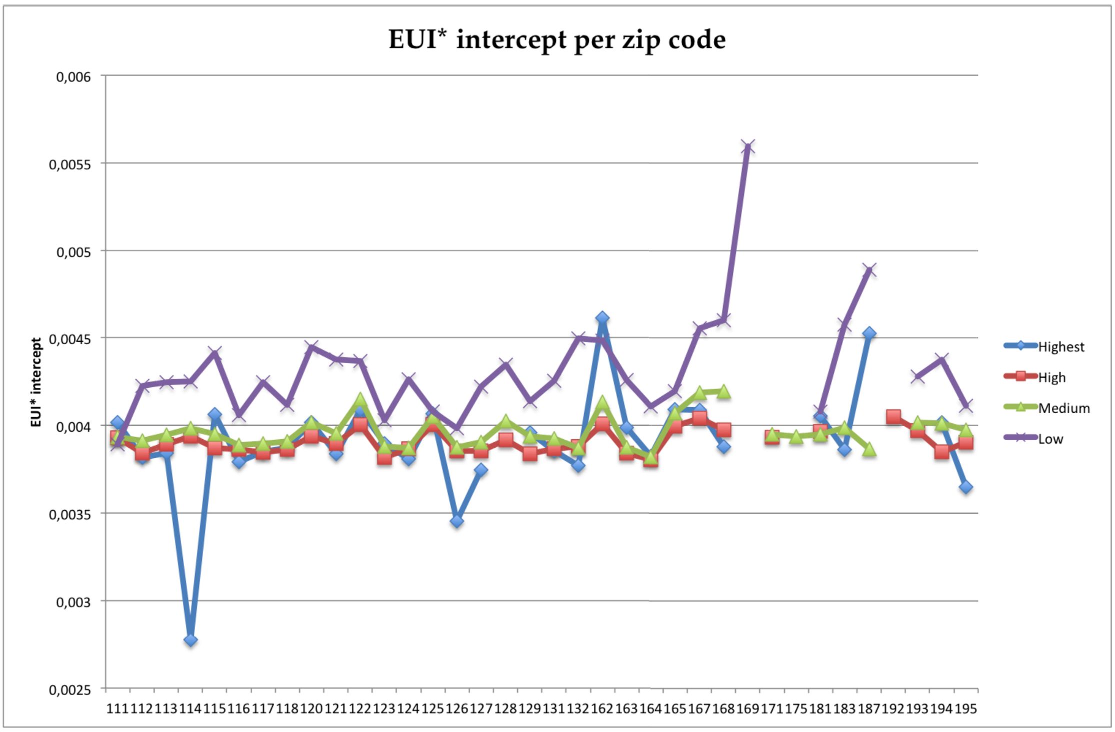


FIGURE 4.40: Comparison of EUI* intercepts for different consumption groups per zip code

The comparison of the temperature and wind dependencies is plotted in figure 4.41 and 4.42 respectively. The temperature dependence graph shows the absolute value (negative) of the temperature regression coefficient. Thus, a higher value on the graph indicates a higher dependency.

Apparently, interdependence of temperature and EUI* is the highest for the low consumption group. This result must be interpreted carefully, since the R squared values for low consumption buildings are low. Consequently, the linear relation for medium and high consumption buildings is increasingly lower. Again, the course of the coefficients for the highest consumption group has a higher variability. Sometimes, buildings that have a high consumption intensity are depending more on the outside temperature, but in general, the dependence becomes less important with higher EUI.

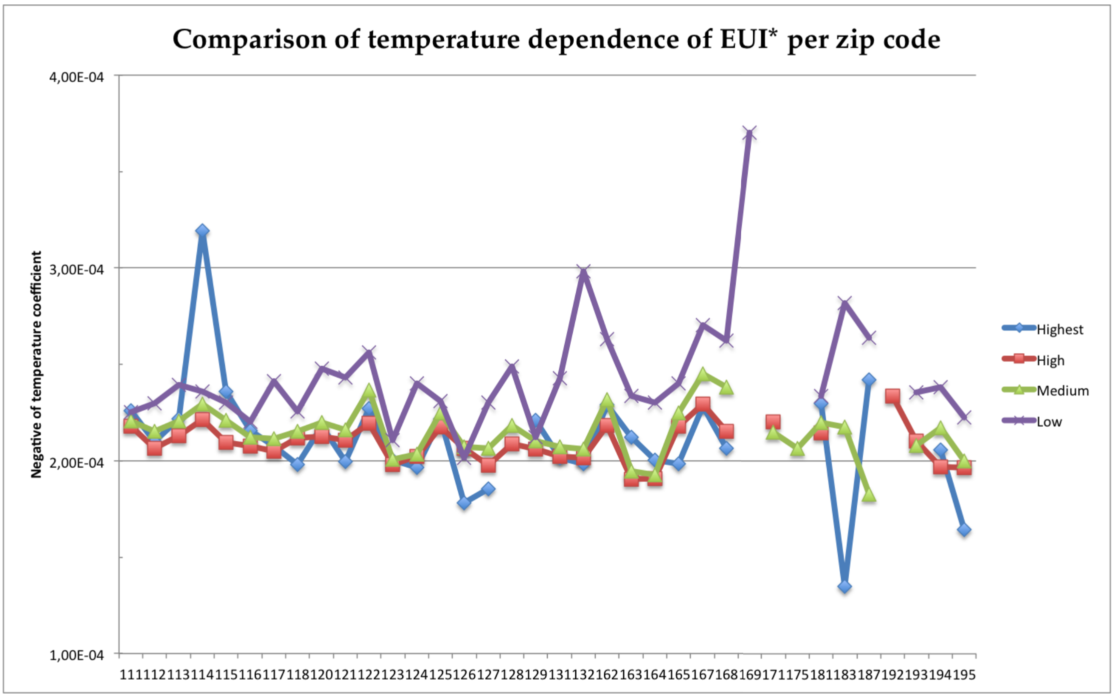


FIGURE 4.41: Comparison of temperature influence for different consumption groups per zip code

The scale of the vertical axis for the wind speed plot (figure 4.42) is chosen with a focus on the stabler medium and high consumption group coefficients. The extremely high and low values in the two other consumption groups are esteemed less important for the sake of the analysis.
Contrary to the temperature dependence, the wind speed dependence is slightly less for the medium consumption group than for the high consumers. The course of the coefficients for the highest consumption group is inconclusive – sometimes higher than the low and medium group, sometimes lower. Regardless of the trustworthiness of the low consumption coefficients, they seem to be generally lower than the other coefficients. This contrasting behaviour for the two coefficients must be investigated closer.

For the temperature and wind coefficients, a slightly decreasing trend from the low zip codes to the higher numbers is apparent. Since the lower postal numbers are located in the city center, this again indicates a higher weather influence in the city center than in the suburbs.

4.7.3 Regression grouped by building category

In a final regression analysis, the spatial aspect is briefly omitted, in order to focus on weather influences in different building types. As seen in table 3.1, there are five categories within the residential building class, viz.

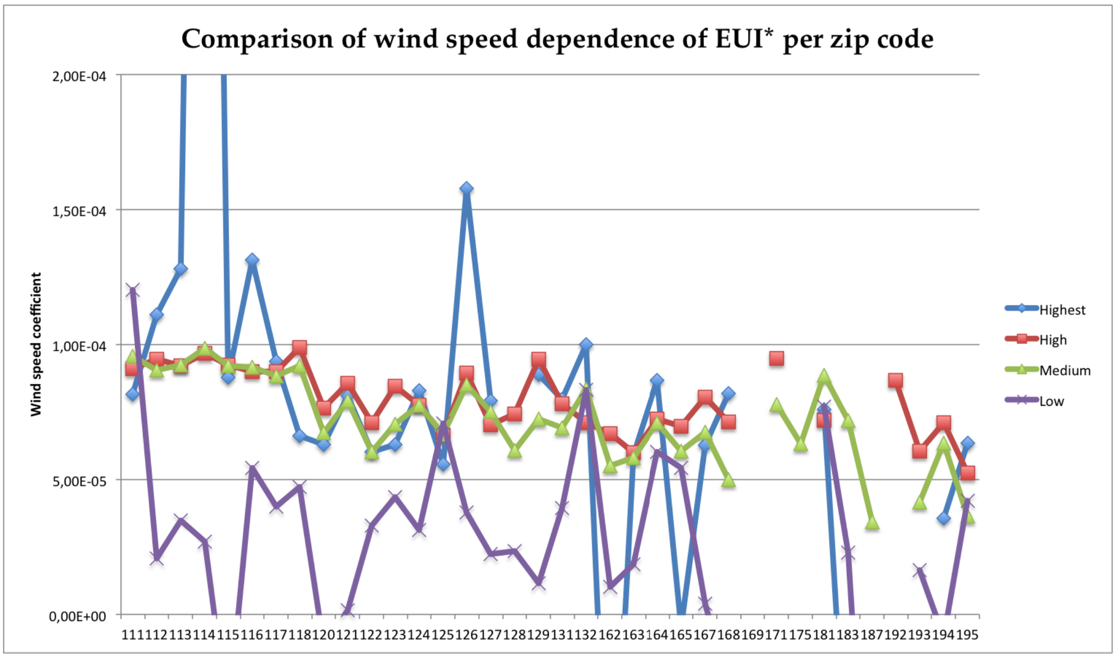


FIGURE 4.42: Comparison of wind speed influence for different consumption groups per zip code

- Single-family & detached house,
- Multi-family house with additional function,
- Apartment in detached multi-family house,
- Detached multi-family house and
- Multi-family house in building block.

Table 4.8 summarizes the results for the regression analysis per residential building group. It can be seen that the intercept is approximately $4 \cdot 10^{-3}$ again, the temperature coefficient lies around $-2.2 \cdot 10^{-4}$ and that the wind speed coefficient fluctuates between $4.55 \cdot 10^{-5}$ and $8.51 \cdot 10^{-5}$. Although the P values indicate undoubtedly that these coefficients are significantly different from 0, the R squared values are on the low sides. This can in part be explained by the high number of observations that have been regressed, bringing along a higher variability and consequently a larger prediction error.

Although the table is not too elaborate, a parallel coordinates chart can help visualising patterns. Such a chart is presented in figure 4.43. The values of the regression coefficients are normalised with respect to the lowest (0) and highest (1) observation. These respective values can be looked up in table 4.8.

In this way, it can be observed that the single-family homes have by far the largest intercept value and temperature influence (the minimal value for the temperature influence

Coefficient	Coeff.	Std. Err.	t-value	P-value	R squared
Detached multi-family house					
Temperature	-2,12E-04	1,61E-07	-1311,97	0,00	0,55
Wind speed	6,83E-05	1,03E-06	66,02	0,00	0,55
Intercept	3,96E-03	3,59E-06	1102,80	0,00	0,55
Multi-family house in building block					
Temperature	-2,17E-04	1,40E-07	-1550,72	0,00	0,59
Wind speed	8,51E-05	8,97E-07	94,94	0,00	0,59
Intercept	3,94E-03	3,11E-06	1266,94	0,00	0,59
Multi-family house with additional function					
Temperature	-2,27E-04	2,37E-07	-956,84	0,00	0,87
Wind speed	7,60E-05	1,52E-06	50,03	0,00	0,87
Intercept	4,04E-03	5,27E-06	766,09	0,00	0,87
Apartment in detached multi-family house					
Temperature	-2,05E-04	1,15E-06	-178,17	0,00	0,36
Wind speed	4,55E-05	7,38E-06	6,17	0,00	0,36
Intercept	3,98E-03	2,56E-05	155,59	0,00	0,36
Single-family & detached house					
Temperature	-2,44E-04	2,04E-07	-1195,59	0,00	0,62
Wind speed	5,42E-05	1,31E-06	41,48	0,00	0,62
Intercept	4,22E-03	4,54E-06	930,26	0,00	0,62

TABLE 4.8: Results of regression grouped by building category

corresponds with the highest influence), but a low influence of wind speed. The multi-family homes with additional function are positioned in the intermediate region for all coefficients. The three remaining building types have more or less the same intercept value and display the same low temperature influence. However the wind influence is very different for these three types: while the apartments inside a larger building experience the smallest effect from wind, the detached multi-family homes and multi-family homes in a block of buildings are affected increasingly.

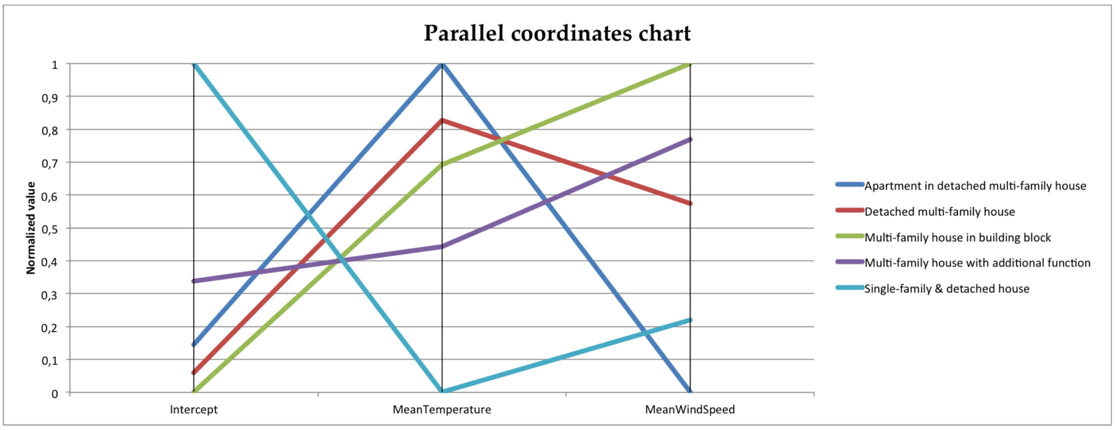


FIGURE 4.43: Parallel coordinate chart of the regression coefficients per building type